

Conditional tests for elliptical symmetry using robust estimators

Ana M. Bianco

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET

Graciela Boente

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET

Isabel M. Rodrigues

Departamento de Matemática and CEMAT, Instituto Superior Técnico, Universidade de Lisboa

Abstract

This paper presents a procedure for testing the hypothesis that the underlying distribution of the data is elliptical when using robust location and scatter estimators instead of the sample mean and covariance matrix. Under mild assumptions that include elliptical distributions without first moments, we derive the test statistic asymptotic behaviour under the null hypothesis and under special alternatives. Numerical experiments allow to compare the behaviour of the tests based on the sample mean and covariance matrix with that based on robust estimators, under various elliptical distributions and different alternatives. This comparison was done looking not only at the observed level and power but we rather use the size-corrected relative exact power which provides a tool to assess the test statistic skill to detect alternatives. We also provide a numerical comparison with other competing tests.

Key Words: Bootstrap, Conditional test, Elliptical symmetry, Robust estimation

AMS Subject Classification: MSC 62H15; 62F35; 62F40

1 Introduction

The family of elliptically symmetric (or elliptically contoured) distributions generalizes the family of multivariate normal distributions. One advantage of the elliptical distributions is that they define a much broader class of multivariate distributions than the multivariate normal distributions so that they can serve as the basis for the development of more robust analyses. In fact, in many situations, normal-theory analyses can be modified slightly retaining their validity across all elliptical distributions. The fact that many statistical procedures (including principal component analysis) yield superior performance when data support elliptical symmetry, motivates the consideration of testing for elliptical symmetry, instead of testing for other forms of multivariate symmetry.

Zhu and Neuhaus (2003) introduced conditional test procedures for testing elliptical symmetry of a multivariate distribution. The conditional tests are exactly valid if the center and the shape matrix are known and are asymptotically valid if they are estimated, when fourth moments exist. It is worth noting that the test proposed by Zhu and Neuhaus (2003) are based on the sample mean and the sample covariance matrix, when the center and/or the shape parameters are unknown. This entails that the test statistics are asymptotically valid only for elliptical distributions such that $\mathbb{E}(\|\mathbf{X}\|^4) < \infty$. In a robust framework, one frequently assumes that the sample belongs to a neighbourhood of a given central elliptical distribution P_0 . The distributions P to be considered in the neighbourhood include heavy tailed distributions. Furthermore, in order to ensure Fisher-consistency of the proposed estimators, it is generally assumed that the resulting distribution P is also elliptical. So, it is of interest to check if the assumption of elliptical symmetry is valid without making moment assumptions. For this reason, in this paper, we propose a testing procedure that can be helpful to decide if a given sample has a common elliptical distribution without requiring moment conditions when consistent estimators of the unknown parameters are available.

The paper is organized as follows. In Section 2, we introduce our proposal, while asymptotic distribution results under the null and under contiguous alternatives are provided in Section 3. A bootstrap method to compute effectively an approximation of the proposed test is presented in Section 4. The results of a Monte Carlo study in dimensions $p = 2$ and 5 are summarized in Section 6, while a procedure to compute the test statistic is described in Appendix A. Besides, in Section 5, we analyse the behaviour of the proposed test statistic and the classical one under different distributions and sample sizes, so as to check their ability to reject the null hypothesis against a

set of alternatives. Proofs are relegated to Appendix B and C.

2 Test statistic

For the sake of completeness, we briefly recall the notion of elliptical symmetry. One can define symmetry in terms of structural properties of the distribution function, of the density function or of the characteristic function. The distribution of a p -dimensional random vector \mathbf{Y} is called spherically symmetric when, for every $p \times p$ matrix $\mathbf{A} \in \mathcal{O}(p)$ (the orthogonal group), the distribution of \mathbf{AY} is the same as that of \mathbf{X} . A random vector $\mathbf{X} \in \mathbb{R}^p$ has an elliptically symmetric or elliptically contoured distribution, with parameters $\boldsymbol{\mu} \in \mathbb{R}^p$ and a non-singular matrix $\boldsymbol{\Sigma}$, if $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ is a spherically symmetric random vector. If this elliptical distribution has finite second moments, then $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is up to a scalar the covariance matrix. More generally, under no moment conditions, the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are called the location and the scatter matrix parameters, respectively. The associated characteristic function of an elliptical vector has the form $\phi(\mathbf{t}) = e^{i\mathbf{t}^T \boldsymbol{\mu}} \psi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$, for $\mathbf{t} \in \mathbb{R}^p$, for some scalar function $\psi : \mathbb{R} \rightarrow \mathbb{R}$. Then, if second moment exists $\text{VAR}(\mathbf{X}) = -2\psi'(0)\boldsymbol{\Sigma}$, where for identifiability of $\boldsymbol{\Sigma}$ it is usually required that $\psi'(0) = -1/2$. We will write that $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. If $\psi(x) = e^{-x/2}$ in the expression of $\phi(\mathbf{t})$, we get the characteristic function of a normal distribution, so elliptical distributions are generalizations of normal distributions. For an overview on these distributions we refer to Fang and Anderson (1990) and Fang *et al.* (1990).

Among the tests for spherical and elliptical distributions that have been introduced, we can mention Beran (1979), Tyler (1982), Baringhaus (1991), Fang *et al.* (1993), Koltchinskii and Li (1998), Koltchinskii and Sakhanenko (2000), Schott (2002), Zhu and Neuhaus (2003), Huffer and Park (2007) and more recently, Batsidis and Zografos (2013) and Batsidis *et al.* (2014).

The goal of this section is to suggest a modification of the conditional test for ellipsoidal symmetric multivariate distributions proposed in Zhu and Neuhaus (2003), which allows its application to data coming from heavy tailed elliptical distributions. This is particularly appealing in a robust framework, since many resistant statistical procedures assume that the underlying distribution is elliptical to get Fisher-consistent estimators.

Let \mathbf{X} be a p -dimensional random vector with distribution P . Given independent and identically distributed (i.i.d.) observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ such that $\mathbf{X}_i \sim \mathbf{X}$, denote by P_n the empirical measure

based on the sample points. Moreover, let $P_n f = P_n(f)$ stand for $(1/n) \sum_{j=1}^n f(\mathbf{X}_j)$ for any function $f : \mathbb{R}^p \rightarrow \mathbb{R}$.

From now on, denote \mathcal{E}_p the class of all elliptical contoured distributions. The hypothesis to be tested is $H_0 : P \in \mathcal{E}_p$, that is, $H_0 : \mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. For each fixed \mathbf{b} and \mathbf{A} define the functional

$$T_{\mathbf{b}, \mathbf{A}}(P) = \int_{\mathcal{S}_p} \int \left[\mathbb{E}_P \left(\sin \left\{ t \mathbf{a}^T \mathbf{A}^{-1/2} (\mathbf{X} - \mathbf{b}) \right\} \right) \right]^2 w(t) dt dv(\mathbf{a}) \quad (1)$$

where $w : \mathbb{R} \rightarrow \mathbb{R}$ is a weight function, $\mathcal{S}_p = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1\}$, v the uniform distribution on \mathcal{S}_p and \mathbb{E}_P indicates that the expectation is taken with respect to the probability measure P . Note that if $\mathbf{X} \sim P = \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, then $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ is spherically distributed. Hence, the imaginary part of its characteristic function vanishes, that is, $\mathbb{E}_P [\sin(t \mathbf{a}^T \mathbf{Z})] = 0$ for any $t \in \mathbb{R}$ and $\mathbf{a} \in \mathcal{S}_p$ which implies that $T_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(P) = 0$.

When $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known, the empirical version of $T_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(P)$ will also be close to zero. This suggests to reject the null hypothesis H_0 for large values of the test statistic $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$T_n(\mathbf{b}, \mathbf{A}) = \int_{\mathcal{S}_p} \int \left\{ \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{A}^{-1/2} (\mathbf{X} - \mathbf{b}) \right] \right\}^2 w(t) dt dv(\mathbf{a}). \quad (2)$$

The statistic $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ was considered by Zhu and Neuhaus (2003) when $w : \mathbb{R} \rightarrow \mathbb{R}$ has a compact support \mathcal{I} and is a weighted version of that studied by Ghosh and Ruymgaart (1992).

Usually the location and scatter matrix parameters are unknown. To overcome this problem, one may replace in $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by consistent estimators. Zhu and Neuhaus (2003) suggested to use the classical (CL) sample estimators, leading to the test statistic $T_{n, \text{CL}} = T_n(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \sum_{i=1}^n \mathbf{X}_i / n$ and $\hat{\boldsymbol{\Sigma}} = \mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T / n$. As mentioned in Anderson *et al.* (1986), for a sub-class of elliptical distributions, the maximum likelihood estimator of $\boldsymbol{\mu}$ is $\bar{\mathbf{X}}$, while that of $\boldsymbol{\Sigma}$ is a constant multiple (depending on the family) of the sample covariance matrix, that is, the estimators have the same form as in the normal case, which justifies the above choice. Anderson *et al.* (1986) studied the general situation of elliptical random matrices $\mathbb{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ which includes the setting of independent columns we are considering. In the particular case of independent random vectors \mathbf{X}_i , the class of elliptical distributions for which the sample mean and covariance matrix (except for a constant) are still the maximum likelihood estimators includes the situation when the density of \mathbf{X}_1 equals $\det(\boldsymbol{\Sigma})^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))$ for some $g : \mathbb{R} \rightarrow \mathbb{R}^+$,

where $y^{p/2}g(y)$ has a finite positive maximum. A condition ensuring the existence of finite positive maximum is the continuity of g and that $\mathbb{E} \|\mathbf{X}\|^2 < \infty$ (see Lemma 2 in Anderson *et al.*, 1986). On the other hand, if the underlying distribution is heavy tailed the values of these estimators may be distorted, rendering meaningless the test results. A solution to this problem is well known in robust statistics: $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have to be estimated in a robust manner, to provide consistent estimators even if moments do not exist as in the case of a multivariate Cauchy distribution.

The proposal in this paper, consists in plugging into the conditional test statistic $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ robust consistent estimators \mathbf{m}_n and \mathbf{V}_n of the location $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$, respectively, to test $H_0 : P \in \mathcal{E}_p$. This leads to the following conditional robust based statistic

$$T_{n,\mathbf{m},\mathbf{V}} = T_n(\mathbf{m}_n, \mathbf{V}_n) = \int_{\mathcal{S}_p} \int \left\{ \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] \right\}^2 w(t) dt dv(\mathbf{a}),$$

where P_n , \mathbf{a} , w and v are defined as in (2). Denote $\mathbf{V}(P)$ and $\mathbf{m}(P)$ the functionals related to \mathbf{V}_n and \mathbf{m}_n , respectively, when $\mathbf{X} \sim P$. Usually, under H_0 , $\mathbf{m}(P) = \boldsymbol{\mu}$ and $\mathbf{V}(P)$ is, up to a multiplicative constant, equal to $\boldsymbol{\Sigma}$. Then, the functional related to $T_n(\mathbf{m}_n, \mathbf{V}_n)$ is just $T_{\mathbf{m}(P), \mathbf{V}(P)}(P)$ defined in (1) which justifies the considered procedure. In Appendix 6.4, we describe a numerical procedure to compute this test statistic.

Remark 2.1. The results on characteristic functions given in Ushakov (1999) give some insight with respect to the choice of the weight function w . Indeed, as defined in Ushakov (1999) a characteristic function Φ is said to be analytic if there exists a function $g : \mathbb{C} \rightarrow \mathbb{C}$ which is analytic in $\{|z| \leq R\}$ for some $R > 0$ and such that $\Phi(t) = g(t)$ for any $t \in [-R, R]$. Theorem 1.7.7 in Ushakov (1999) states that if a characteristic function φ coincides with an analytic characteristic function Φ in some real neighbourhood of the origin, then they coincide for all real, that is, $\varphi = \Phi$.

Given a random vector $\mathbf{X} \in \mathbb{R}^p$, denote $Z_{\mathbf{a}}$ the random variable $Z_{\mathbf{a}} = \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})$ and $\varphi_{\mathbf{a}}$ its characteristic function. Assume that for any $\mathbf{a} \in \mathcal{S}_p$, $\varphi_{\mathbf{a}}$ is an analytic characteristic function. Theorem 1.7.7 in Ushakov (1999) entails that if, for any $\mathbf{a} \in \mathcal{S}_p$, and for some $\delta > 0$ we have that $\varphi_{\mathbf{a}}(t) = \text{Re}(\varphi_{\mathbf{a}})(t)$ for $t \in (-\delta, \delta)$ then, the random variable $Z_{\mathbf{a}}$ has a symmetric distribution for all \mathbf{a} . So, in the situation where the underlying distribution is such that all the projections have analytic characteristic functions, which includes the multivariate normal as well as the uniform distribution on the ball or in sphere, if the functional related to the test statistic is zero for some weight function w , with support around 0, then it will be 0 for any weight function. Thus, for

probability measures such that for all $\mathbf{a} \in \mathcal{S}_p$, the distribution of $Z_{\mathbf{a}}$ has an analytic characteristic function the choice of w is not crucial as far as its support contains a neighbourhood of 0.

3 Asymptotic behaviour of the test statistic

In order to derive the limit behaviour of the proposed test statistic, we will assume that $w(t)$ has bounded support contained in some finite interval \mathcal{I} and we will introduce the empirical process given by $\mathcal{W}_n = \left\{ \mathcal{W}_n(t, \mathbf{a}) = \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] , (t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p \right\}$. Theorem 3.1 states the asymptotic distribution of the process $\{\mathcal{W}_n(t, \mathbf{a})\}$ under the null hypothesis, while Theorem 3.2 concerns the behaviour under local alternatives. Note that in our statement we do not require neither finite moment conditions to the random vector \mathbf{X} , nor a rate of convergence of the shape matrix estimator. In this sense, our result provides an improvement over the proposal given in Zhu and Neuhaus (2003) who required finite fourth moment. As shown in Section 6, the lack of moments may distort the results of the classical test based on $T_{n,CL}$, while when using robust estimators the test is still reliable. On the other hand, when second moments exist, if we take \mathbf{m}_n and \mathbf{V}_n as the sample mean and covariance matrix, respectively, Theorem 3.1 provides the asymptotic distribution of \mathcal{W}_n under slightly more general conditions than those given in Theorem 2.1 of Zhu and Neuhaus (2003).

3.1 Behaviour under the null hypothesis

Theorem 3.1. *Let \mathcal{I} be a bounded interval. Assume that $\mathbf{X} \sim P = \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, i.e., that H_0 holds and that $\int_0^1 \sqrt{\log H_u} du < \infty$, where H_δ is the smallest value $H \geq 1$ such that $\mathbb{P}(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\| > H) \leq \delta^2$, that is $H_\delta = \max(1, F_S^{-1}(1 - \delta^2))$ with F_S the distribution function of $S = \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|$. Moreover, assume that $\mathbb{P}(\mathbf{X} = \boldsymbol{\mu}) = 0$ and that \mathbf{V}_n and \mathbf{m}_n are consistent estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$, respectively, such that \mathbf{m}_n admits, for some function $\alpha_{\mathbf{m}} : \mathbb{R} \rightarrow \mathbb{R}$, a Bahadur expansion as follows*

$$\sqrt{n}(\mathbf{m}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \alpha_{\mathbf{m}} \left(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\| \right) + o_{\mathbb{P}}(1) , \quad (3)$$

where $\mathbb{E}_P \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|^2 \alpha_{\mathbf{m}}^2(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|) < \infty$.

Then, the process $\mathcal{W}_n = \left\{ \mathcal{W}_n(t, \mathbf{a}) = \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] , (t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p \right\}$ converges in distribution to a centered continuous Gaussian process $\mathcal{W} = \{\mathcal{W}(t, \mathbf{a}) , (t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p\}$ with

covariance kernel given by $\mathbb{E}_P[k(t, \mathbf{a}, \mathbf{X})k(s, \mathbf{b}, \mathbf{X})]$, for $(t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p$ and $(s, \mathbf{b}) \in \mathcal{I} \times \mathcal{S}_p$, where $k(t, \mathbf{a}, \mathbf{x}) = \sin \left[t \mathbf{a}^\top \Sigma^{-1/2} (\mathbf{x} - \boldsymbol{\mu}) \right] - \psi(t^2) t \mathbf{a}^\top \Sigma^{-1/2} (\mathbf{x} - \boldsymbol{\mu}) \alpha_{\mathbf{m}} \left(\|\Sigma^{-1/2} (\mathbf{x} - \boldsymbol{\mu})\| \right)$.

Remark 3.1. Note that the classical location estimator, that is, the sample mean corresponds to $\alpha_{\mathbf{m}}(t) = 1$ and this is the situation considered in Theorem 2.1 of Zhu and Neuhaus (2003) which requires the existence of fourth moments. On the other hand, as shown in Hampel *et al.* (1986), if \mathbf{m}_n is an estimator related to a functional $\mathbf{m}(P)$ that is affine equivariant there exists a real function $\alpha_{\mathbf{m}} : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that its influence function equals $\text{IF}(\mathbf{x}_0, \mathbf{m}, P_0) = (\mathbf{x}_0 - \boldsymbol{\mu}) \alpha_{\mathbf{m}}(\|\Sigma^{-1/2}(\mathbf{x}_0 - \boldsymbol{\mu})\|)$. In most cases, the influence function is bounded so the assumption $\mathbb{E}_P \|\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|^2 \alpha_{\mathbf{m}}^2(\|\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\|) < \infty$ is satisfied and no moment conditions are required. Besides, as it is well known, under mild conditions, the influence function allows to obtain a Bahadur expansion for the location estimator (see Fernholz, 1983). In particular, for the S -estimator (see Lopuhaä, 1989), we have that $\alpha_{\mathbf{m}}(t) = \beta^{-1} u_s(t)$ where $u_s(t) = \psi_s(t)/t$, $\psi_s(t) = \rho'_s(t)$. Let G_0 be the spherical distribution related to P_0 , that is, G_0 is the distribution of $\Sigma^{-1/2}(\mathbf{x}_0 - \boldsymbol{\mu})$. Then, the constant β is given by

$$\beta = \mathbb{E}_{G_0} \left[\left(1 - \frac{1}{p} \right) u_s(\|\mathbf{Z}\|) + \frac{1}{p} \psi'_s(\|\mathbf{Z}\|) \right]. \quad (4)$$

Usually, the influence function is computed at the central Gaussian distribution, so that $G_0 = N(\mathbf{0}, \mathbf{I})$. A common choice for the ρ -function defining the S -estimator is the Tukey function defined as $\rho_s(y) = (c^2/6) \min\{1 - [1 - (y/c)^2]^3, 1\}$. Hence, $\psi_s(y) = y [1 - (y^2/c^2)]^2 \mathbb{I}_{[-c, c]}(y)$, $\psi'_s(y) = [1 - 6(y^2/c^2) + 5(y^4/c^4)] \mathbb{I}_{[-c, c]}(y)$ and $u_s(y) = [1 - (y^2/c^2)]^2 \mathbb{I}_{[-c, c]}(y)$ (see Lopuhaä, 1989).

Remark 3.2. Let us show that the assumption $\int_0^1 \sqrt{\log H_u} du < \infty$ where $\mathbb{P}(\|\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\| > H_\delta) \leq \delta^2$, is fulfilled for some distributions where fourth moments may not exist. For the sake of simplicity, we will assume $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \mathbf{I}_p$, since otherwise, we may consider $\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$.

It is clear that if $\mathbb{E}_P \|\mathbf{Z}\|^2 < \infty$, then $H_\delta \leq (\mathbb{E}_P \|\mathbf{Z}\|^2)^{1/2} / \delta$ and $\int_0^1 \sqrt{\log H_u} du < \infty$. More generally, if $\mathbb{E}_P \|\mathbf{Z}\|^\nu < \infty$, for some $\nu > 0$, then $H_\delta \leq (\mathbb{E}_P \|\mathbf{Z}\|^\nu)^{1/\nu} / \delta^{1/\nu}$, so $\int_0^1 \sqrt{\log H_u} du < \infty$.

As an example of elliptical distributions satisfying the condition $\int_0^1 \sqrt{\log H_u} du < \infty$, let us consider the multivariate t -distribution with k degrees of freedom, i.e., $\mathbf{Z} \sim \mathcal{T}_{p,k}(\mathbf{0}, \mathbf{I})$. As is well known \mathbf{Z} has no finite fourth moment when $k \leq 4$. Besides, \mathbf{Z} has the same distribution as $y^{1/2} \mathbf{W}$

where $v_k = ky^{-1} \sim \chi_k^2$ and $\mathbf{W} \sim N_p(\mathbf{0}, \mathbf{I})$, where χ_k^2 stands for the chi-square distribution with k degrees of freedom. Then, if the two expectations on the right hand side of (5) below exist, by the independence between \mathbf{W} and y we have that

$$\mathbb{E}\|y^{1/2}\mathbf{W}\|^\nu = k^m \mathbb{E}\left(\frac{\|\mathbf{W}\|^\nu}{v_k^m}\right) = k^m \mathbb{E}\|\mathbf{W}\|^\nu \mathbb{E}(v_k^{-m}), \quad (5)$$

where $m = \nu/2$. Note that $\mathbb{E}\|\mathbf{W}\|^\nu < \infty$ for any $\nu > 0$. On the other hand, using that $v_k \sim \chi_k^2$, we have that, for any $0 < \nu < k$, $\mathbb{E}(v_k^{-\nu/2}) < \infty$, which entails that for the multivariate t -distribution $\int_0^1 \sqrt{\log H_u} du < \infty$. This result shows that our assumption is a very mild one since it includes for instance, the multivariate Cauchy distribution.

The following Corollary gives the distribution of the test statistic under the null hypothesis.

Corollary 3.1. *Under the assumptions of Theorem 3.1 if $w(t)$ is a weight function with bounded support \mathcal{I} , then $T_n(\mathbf{m}_n, \mathbf{V}_n) \xrightarrow{D} \int_{\mathcal{S}_p} \int_{\mathcal{I}} \mathcal{W}^2(t, \mathbf{a}) w(t) dt dv(\mathbf{a})$ where the process $\mathcal{W}(t, \mathbf{a})$ is defined in Theorem 3.1.*

3.2 Behaviour under the alternative

Regarding the consistency of the test, it is well known that if $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, then for any $\mathbf{a} \in \mathcal{S}_p$, $Z_{\mathbf{a}} = \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has a symmetric distribution, but the converse is not true. A typical example being a random variable \mathbf{Y} with distribution uniform on the set $\{\mathbf{x} : -1 \leq x_j \leq 1 \text{ for all } j\}$ which satisfies that $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(-\mathbf{y})$ ensuring that all projections $\mathbf{a}^T \mathbf{Y}$ are symmetric. In this situation, for observations having a symmetric distribution but not a spherical one, the test type-statistics considered in Zhu and Neuhaus (2003), Ghosh and Ruymgaart (1992) or in this paper will not reject the null hypothesis. This is a feature of any projection-pursuit procedure based on the property that any projection of a spherical distributed random vector is symmetric.

On the other hand, as mentioned in Zhu and Neuhaus (2003), if $\mathbb{E}_P \sin(t\mathbf{a}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})) \neq 0$, for some $t \in \mathcal{I}$ and $\mathbf{a} \in \mathcal{S}_p$, then $T_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(P) > 0$. Therefore, using that $T_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})/n \xrightarrow{p} T_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(P)$ together with the consistency of \mathbf{m}_n and \mathbf{V}_n , we obtain that the test statistic $T_n(\mathbf{m}_n, \mathbf{V}_n)$ will converge to ∞ and the test is consistent against *global alternatives*.

To derive the distribution of the test statistic under a set of alternatives, denote as $\sin^{(j)}(t)$ the

j -th derivative of the sinus function at t . Recall that $\mathbf{V}(P)$ and $\mathbf{m}(P)$ stand for the functionals related to \mathbf{V}_n and \mathbf{m}_n , respectively, when $\mathbf{X} \sim P$. We will assume that $\mathbf{m}(P)$ is affine equivariant. To strength the dependence on the sample, we will denote as $P_{n,\mathbf{X}}$ the empirical distribution of the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ and $\mathbf{V}_{n,\mathbf{X}}$ and $\mathbf{m}_{n,\mathbf{X}}$ the estimators based on that sample, that is, $\mathbf{V}_{n,\mathbf{X}} = \mathbf{V}(P_{n,\mathbf{X}})$ and $\mathbf{m}_{n,\mathbf{X}} = \mathbf{m}(P_{n,\mathbf{X}})$. Moreover, assume that the i.i.d. observation $\mathbf{X}_i = \mathbf{X}_{in}$ are such that $\mathbf{X}_i = \mathbf{X}_{in} = \mathbf{Z}_i + \mathbf{Y}_i n^{-\alpha}$ for some $\alpha > 0$ where \mathbf{Z}_i are i.i.d. such that $\mathbf{Z}_i \sim P_0 = \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. Due to the equivariance of the location estimator and without loss of generality, we may assume that $\mathbb{E}(\mathbf{Y}) = \mathbf{0}$. Effectively, if we define $\tilde{\mathbf{X}}_i = \mathbf{X}_i - \mathbb{E}(\mathbf{Y})/n^\alpha$, using that $\mathbf{m}_{n,\tilde{\mathbf{X}}} = \mathbf{m}_{n,\mathbf{X}} - \mathbb{E}(\mathbf{Y})/n^\alpha$, we have that $P_n \sin \left[t \mathbf{a}^\top \mathbf{V}_n^{-1/2} (\tilde{\mathbf{X}} - \mathbf{m}_{n,\tilde{\mathbf{X}}}) \right] = P_n \sin \left[t \mathbf{a}^\top \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_{n,\mathbf{X}}) \right]$, so to obtain the asymptotic behaviour of the test statistic under these alternatives we may assume that $\mathbb{E}(\mathbf{Y}) = \mathbf{0}$.

Theorem 3.2. *Let $w(t)$ be a weight function with support \mathcal{I} and $\mathbf{Z} \sim P_0 = \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. Assume that the following assumptions hold*

- a) \mathbf{Z}_i are i.i.d. such that $\mathbf{Z}_i \sim \mathbf{Z}$ and define $\mathbf{X}_i = \mathbf{X}_{in} = \mathbf{Z}_i + \mathbf{Y}_i n^{-\alpha}$, where $\alpha > 0$ and $\mathbb{E}(\mathbf{Y}) = \mathbf{0}$.
- b) $\mathbb{P}(\mathbf{Z} = \boldsymbol{\mu}) = \mathbb{P}(\mathbf{X} = \boldsymbol{\mu}) = 0$,
- c) $\int_0^1 \sqrt{\log H_u} du < \infty$, where H_δ is the smallest value $H \geq 1$ such that $\mathbb{P}(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\| > H) \leq \delta^2$, that is $H_\delta = \max(1, F_S^{-1}(1 - \delta^2))$ where F_S is the distribution function of $S = \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\|$.
- d) $\mathbf{V}_{n,\mathbf{X}}$ and $\mathbf{m}_{n,\mathbf{X}}$ are such that $\mathbf{V}_{n,\mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$ and $\mathbf{m}_{n,\mathbf{X}} - \boldsymbol{\mu} \xrightarrow{p} 0$.
- e) The functional $\mathbf{m}(P)$ is such that the Bahadur expansion (3) holds at P_0 , that is, for some function $\alpha_{\mathbf{m}} : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{m}_{n,\mathbf{Z}}$ satisfies (3) where $\mathbb{E}_{P_0} \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\|^2 \alpha_{\mathbf{m}}^2(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\|) < \infty$. Furthermore,

$$\sqrt{n}(\mathbf{m}_{n,\mathbf{X}} - \boldsymbol{\mu}) = \sqrt{n}(\mathbf{m}_{n,\mathbf{Z}} - \boldsymbol{\mu}) + o_{\mathbb{P}}(1). \quad (6)$$

- f) there is a positive integer m such that $\mathbb{E}\|\mathbf{Y}\|^{2m} < \infty$ and for some $\delta_m > 0$

$$\sup_{(t,\mathbf{a}) \in \mathcal{I}^{\delta_m} \times \mathcal{S}_p} |B_m(t, \mathbf{a})| = \sup_{(t,\mathbf{a}) \in \mathcal{I}^{\delta_m} \times \mathcal{S}_p} \left| \mathbb{E} \left\{ \left(t \mathbf{a}^\top \boldsymbol{\Sigma}^{-1/2} \mathbf{Y} \right)^m \sin^{(m)}(t \mathbf{a}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})) \right\} \right| \neq 0,$$

where \mathcal{I}^δ stands for a δ -neighborhood of $\mathcal{I} = [-\nu, \nu]$.

Let ℓ be the smallest positive m satisfying f). If \mathbf{Y} and \mathbf{Z} are not independent assume in addition that for any $1 \leq s < \ell$, $\int_0^1 \sqrt{\log H_{u^q}} du < \infty$, where $q = \ell/(\ell - s)$. Then, when $\alpha = 1/(2\ell)$,

$$T_n(\mathbf{m}_n, \mathbf{V}_n) \xrightarrow{D} \int_{\mathcal{S}_p} \int_{\mathcal{I}} \left(\mathcal{W}(t, \mathbf{a}) + \frac{1}{\ell!} B_\ell(t, \mathbf{a}) \right)^2 w(t) dt dv(\mathbf{a}),$$

where the process $\mathcal{W}(t, \mathbf{a})$ is defined in Theorem 3.1.

4 Bootstrap method

As mentioned in Zhu and Neuhaus (2003), the asymptotic behaviour of the test statistic does not allow to compute easily p -values, so a bootstrap method is needed. Zhu and Neuhaus (2003) describe a bootstrap procedure when the center $\boldsymbol{\mu}$ and the shape parameter $\boldsymbol{\Sigma}$ are known and when $\boldsymbol{\mu}$ is estimated using the mean of the observations. When the center and the shape matrix are unknown, a slight modification to the method considered in Zhu and Neuhaus (2003) is needed to adapt to the resistant location estimators, since the estimated shape parameter does not influence the distribution of the test statistic.

One possibility is to adapt the bootstrap statistic defined in Zhu and Neuhaus (2003) to the present setting. For that purpose, assume, as in Section 3, that the affine equivariant location estimator admits a Bahadur expansion given by (3). Let $\hat{\mathbf{d}} = \mathbf{a}^T \mathbf{V}_n^{-1/2}$. Using that

$$\begin{aligned} P_n \sin \left[t \hat{\mathbf{d}}^T (\mathbf{X} - \mathbf{m}_n) \right] &= \cos \left[t \hat{\mathbf{d}}^T (\boldsymbol{\mu} - \mathbf{m}_n) \right] P_n \sin \left[t \hat{\mathbf{d}}^T (\mathbf{X} - \boldsymbol{\mu}) \right] \\ &\quad - \sin \left[t \hat{\mathbf{d}}^T (\mathbf{m}_n - \boldsymbol{\mu}) \right] P_n \cos \left[t \hat{\mathbf{d}}^T (\mathbf{X} - \boldsymbol{\mu}) \right], \end{aligned}$$

the first order von Mises expansion (3) and the fact that $\mathbf{m}_n \xrightarrow{p} \boldsymbol{\mu}$, we have that

$$\sqrt{n} P_n \sin \left[t \hat{\mathbf{d}}^T (\mathbf{X} - \mathbf{m}_n) \right] = \sqrt{n} P_n \sin (t \mathbf{a}^T \mathbf{z}) - \sqrt{n} \sin \{ t \mathbf{a}^T P_n [\mathbf{z} \alpha_{\mathbf{m}}(\|\mathbf{z}\|)] \} P_n [\cos (t \mathbf{a}^T \mathbf{z})] + o_{\mathbb{P}}(1),$$

where $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \sim \mathbf{u} \|\boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu})\|$ and $\mathbf{u} \sim \mathcal{U}(\mathcal{S}_p)$.

However, the implementation of this bootstrapping method when using a location and scatter matrix robust estimators implies the computation of $\alpha_{\mathbf{m}}(t)$, which equals 1 for the sample mean, but may be more complex when using robust estimators. For instance, when using S -estimators the function $\alpha_{\mathbf{m}}(t)$ involves the calculation of the constant β defined in (4). In robustness, β is usually

computed under the standard normal distribution. However, in our situation, the constant β must be computed under the spherical distribution related to the underlying elliptical distribution of the sample. This is a drawback of this bootstrap method since the correct distribution is unknown. To avoid this vicious circle, we consider a bootstrap statistic that can be computed as follows:

- Step 1** Generate i.i.d. random vector $\mathbf{u}_i \sim \mathcal{U}(\mathcal{S}_p)$ and define $\mathbf{U}_n = (\mathbf{u}_1, \dots, \mathbf{u}_n)$. Let $\mathbf{X}_i^* = \mathbf{u}_i \|\mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{m}_n)\|$ be the bootstrap observations and P_n^* the empirical distribution of \mathbf{X}_i^*
- Step 2** Define $W_n^*(t, \mathbf{a}) = \sqrt{n} P_n^* \sin \left[t \mathbf{a}^T \mathbf{V}^{*-1/2} (\mathbf{X}^* - \mathbf{m}^*) \right]$, where \mathbf{m}^* and \mathbf{V}^* are the location and scatter matrix robust estimators of the bootstrapped sample \mathbf{X}_i^* and calculate the statistic $T_{n, \mathbf{m}^*, \mathbf{V}^*}^*(\mathbf{U}_n) = \int_{\mathcal{S}_p} \int W_n^*(t, \mathbf{a})^2 w(t) dt dv(\mathbf{a})$.
- Step 3** Repeat Steps 1 and 2 $Nboot$ times to get $Nboot$ values of $T_{n, \mathbf{m}^*, \mathbf{V}^*}^*(\mathbf{U}_n^{(j)})$, $1 \leq j \leq Nboot$.
- Step 4** Estimate the p -value as $p = k / (Nboot + 1)$ where k is the number of $T_{n, \mathbf{m}^*, \mathbf{V}^*}^*(\mathbf{U}_n^{(j)})$ that are greater or equal than $T_{n, \mathbf{m}, \mathbf{V}}$.

Through this algorithm we obtain a sample of bootstrap replicates T_j^* , $1 \leq j \leq Nboot$ whose distribution approximates the distribution of $T_{n, \mathbf{m}, \mathbf{V}}$ under the null hypothesis, as desired.

The proof of the asymptotic distribution of the bootstrap procedure is an interesting topic which we leave for future research.

5 Finite sample distribution of the test statistic when $p = 2$

We generate independent observations $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, $\mathbf{Z}_i \sim P$ according to different elliptical models under the null hypothesis. Let $\mathcal{T}_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the multivariate p -dimensional t -distribution with k degrees of freedom, which includes the multivariate Cauchy distribution when $k = 1$, and denote $\mathcal{U}(\mathcal{S}_p)$ and $\mathcal{U}(\mathcal{B}_p)$ the uniform distributions over the unit circle and the unit ball, respectively. Consider the null hypotheses $H_0^{(1)} : P = N_p(\mathbf{0}, \mathbf{I})$, $H_0^{(2)} : P = 0.9N_p(\mathbf{0}, \mathbf{I}) + 0.1\mathcal{T}_{p,1}(\mathbf{0}, \mathbf{I})$, $H_0^{(3)} : P = 0.9N_p(\mathbf{0}, \mathbf{I}) + 0.1\mathcal{T}_{p,3}(\mathbf{0}, \mathbf{I})$, $H_0^{(4)} : P = \mathcal{T}_{p,3}(\mathbf{0}, \mathbf{I})$, $H_0^{(5)} : P = \mathcal{U}(\mathcal{S}_p)$, $H_0^{(6)} : P = \mathcal{U}(\mathcal{B}_p)$ and $H_0^{(7)} : P = \mathcal{T}_{p,1}(\mathbf{0}, \mathbf{I})$.

For each null hypothesis $H_0^{(j)}$, we consider different alternative hypothesis $H_{1,\Delta}^{(j)}$, related to the original distribution P in the null hypothesis. Under $H_{1,\Delta}^{(j)}$, the observations are generated as

$\mathbf{X}_i = \mathbf{Z}_i + \Delta \mathbf{Y}_i$ with $\mathbf{Z}_i \sim P$ independent of \mathbf{Y}_i and $\mathbf{Y} = (Y_1, Y_2)^T$ where $Y_k \sim \chi_1^2$ independent among each other and $\Delta = 0.5, 1$ and 1.5 . We also studied the behaviour of the statistics under two fixed alternatives $H_1^{\star(1)}$ and $H_1^{\star(2)}$. Under $H_1^{\star(1)}$, the data have the distribution of a random vector with two independent components, $\mathcal{E}(1)$ and $N(0, 1)$, where $\mathcal{E}(\lambda)$ denotes the exponential distribution of parameter λ , that is, with mean value $1/\lambda$, while the alternative $H_1^{\star(2)}$ corresponds to the distribution of a random vector with two independent components, $\mathcal{E}(1)$ and $\mathcal{E}(1/2)$, that is, with expectation 1 and 2, respectively. The first three alternatives were studied in Zhu and Neuhaus (2003), while Koltchinskii and Li (1998) studied the capability of their proposal to detect $H_1^{\star(1)}$ and $H_1^{\star(2)}$.

In all cases, we perform $N = 1000$ replications for samples of size $n = 20, 50, 100$ and 200 . For each sample, we compute the test statistics with the mean and sample covariance matrix, denoted by $T_{n,CL}$, with the Donoho–Stahel estimators of location and scatter, denoted by $T_{n,DS}$ and with the S –estimators of location and scatter, denoted by $T_{n,S}$. Both robust estimators are calibrated to attain 50% breakdown point. We choose as weight function $w(t) = \mathbb{I}_{[-b,b]}(t)$ with $b = 2$.

In Figures 1 to 7, the density estimates of test statistics $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$ are plotted under the null hypotheses $H_0^{(1)}$ to $H_0^{(7)}$ and under their corresponding alternatives. The density estimates were evaluated using the normal kernel.

As expected, in most cases the classical test statistics is more sensitive to the lack of elliptical symmetry of the alternative distributions than the robust test statistics. However, for $n = 20$ in the considered situations all statistics, the classical and the robust ones, fail to distinguish the symmetric distribution under the null hypothesis from those considered in the alternatives. Indeed, for this sample size all the density estimates are almost overlapping. The tests detect some of the selected alternatives for $n = 50$. For $n = 100$, in all cases the ability of the test statistics to make out the nature of the underlying distribution increases and this fact becomes more clear for $n = 200$. For $n = 100$ and 200 , the densities corresponding to the non–elliptical distributions generated under $H_1^{\star(1)}$ and $H_1^{\star(2)}$ are shifted to the right from those of the test statistic under the null hypothesis. This effect is less visible for $n = 50$. Hence, one could expect that the tests statistics will work well under these circumstances. On the opposite, except for Figure 6, the densities of all the test statistics under the null hypothesis and under the alternative $H_{1,0.5}^{(j)}$ are almost overlapping. As expected, this performance is even worst for the classical test under $H_0^{(7)}$ and $H_{1,\Delta}^{(7)}$, where the distribution of the test statistic does not allow to distinguish between the null and the alternative

hypotheses even for $n = 200$ (see Figure 7). Hence, one can not expect a good performance of the classical tests in this case. A similar conclusion can be held for $H_{1,1}^{(j)}$ for $n = 100$, while for $n = 200$ the behaviour of the test statistic depends on the distribution of \mathbf{Z}_i .

This numerical approach suggests that, for $p = 2$, small sample sizes and values of Δ smaller than 0.5 when considering alternatives of the form $\mathbf{Z}_i + \Delta \mathbf{Y}_i$ should not be considered in the Monte Carlo study presented in Section 6.

6 Monte Carlo study

In this section, we report the results of a simulation study conducted to analyse the performance of the test statistic obtained using robust location and scatter estimators with respect to that based on the sample mean and covariance matrix. The weight function considered in this Monte Carlo study equals $w(t) = \mathbb{I}_{[-b,b]}(t)$ where $b = 2$. Based on the results reported in Section 5 regarding the finite-sample distribution of the test statistics, we carried out $NR = 500$ replications for sample sizes equal to $n = 200$. To perform the bootstrap method described in Section 4, we used $Nboot = 1000$ bootstrap samples. The nominal level was set equal to $\alpha = 0.05$. Besides, we also compare our test procedure with other known methods for testing elliptical symmetry, when $p = 5$. In what follows, $\pi_{H_0}(T_n)$ and $\pi_{H_1}(T_n)$ stand for the observed probabilities of rejection of the test based on the statistic T_n under the null hypothesis and under the alternative H_1 , respectively. As in Batsidis *et al.* (2014), we also examine if the empirical size is significantly different from the nominal level $\alpha = 0.05$. To be more precise, let π be such that $\pi_{H_0}(T_n) \xrightarrow{p} \pi$. Then, using the central limit theorem, the hypothesis $H_{0,\pi} : \pi = \alpha$ is rejected at level γ versus $H_{1,\pi} : \pi \neq \alpha$ if $\pi_{H_0}(T_n) \notin [a_1(\alpha), a_2(\alpha)]$ where $a_j(\alpha) = \alpha + (-1)^j z_{\gamma/2} \{\alpha(1 - \alpha)/NR\}^{1/2}$, $j = 1, 2$. If $H_{0,\pi} : \pi = \alpha = 0.05$ is not rejected, the testing procedure based on T_n is considered accurate. Note that if $\pi_{H_0}(T_n) < a_1(\alpha)$ the testing procedure is conservative, while if $\pi_{H_0}(T_n) > a_2(\alpha)$ the test is liberal. In all Tables reporting the observed frequencies of rejection, we indicate with \star those cases in which the observed empirical frequencies of rejection are different from the nominal level with a significance level $\gamma = 0.01$.

From now on, let $\mathcal{T}_{p,k}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the multivariate p -dimensional t -distribution with k degrees of freedom, which includes the multivariate Cauchy distribution when $k = 1$, and denote $\mathcal{U}(\mathcal{S}_p)$ and $\mathcal{U}(\mathcal{B}_p)$ the uniform distributions over the unit circle and the unit ball, respectively.

Denote also as χ_ν^2 the chi-square distribution with ν degrees of freedom.

6.1 Simulation study in dimension $p = 2$

We generate independent observations $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, $\mathbf{Z}_i \sim P$ according to different elliptical models under the null hypothesis. Consider the null hypotheses $H_0^{(1)} : P = N_p(\mathbf{0}, \mathbf{I})$, $H_0^{(2)} : P = 0.9N_p(\mathbf{0}, \mathbf{I}) + 0.1\mathcal{T}_{p,1}(\mathbf{0}, \mathbf{I})$, $H_0^{(3)} : P = 0.9N_p(\mathbf{0}, \mathbf{I}) + 0.1\mathcal{T}_{p,3}(\mathbf{0}, \mathbf{I})$, $H_0^{(4)} : P = \mathcal{T}_{p,3}(\mathbf{0}, \mathbf{I})$, $H_0^{(5)} : P = \mathcal{U}(\mathcal{S}_p)$, $H_0^{(6)} : P = \mathcal{U}(\mathcal{B}_p)$ and $H_0^{(7)} : P = \mathcal{T}_{p,1}(\mathbf{0}, \mathbf{I})$.

For each null hypothesis $H_0^{(j)}$, we consider different alternative hypothesis $H_{1,\Delta}^{(j)}$, related to the original distribution P in the null hypothesis. Under $H_{1,\Delta}^{(j)}$, the observations are generated as $\mathbf{X}_i = \mathbf{Z}_i + \Delta \mathbf{Y}_i$ with $\mathbf{Z}_i \sim P$ independent of \mathbf{Y}_i and $\mathbf{Y} = (Y_1, Y_2)^T$ where $Y_k \sim \chi_1^2$ independent among each other and $\Delta = 0.5, 1$ and 1.5 . We also studied the behaviour of the statistics under two fixed alternatives $H_1^{\star(1)}$ and $H_1^{\star(2)}$. Under $H_1^{\star(1)}$, the data have the distribution of a random vector with two independent components, $\mathcal{E}(1)$ and $N(0, 1)$, where $\mathcal{E}(\lambda)$ denotes the exponential distribution of parameter λ , that is, with mean value $1/\lambda$, while the alternative $H_1^{\star(2)}$ corresponds to the distribution of a random vector with two independent components, $\mathcal{E}(1)$ and $\mathcal{E}(1/2)$, that is, with expectations 1 and 2, respectively. The first three alternatives were studied in Zhu and Neuhaus (2003), while Koltchinskii and Li (1998) studied the capability of their proposal to detect $H_1^{\star(1)}$ and $H_1^{\star(2)}$.

For each sample, we compute the p -values of the test statistics obtained using the mean and sample covariance matrix, denoted by $T_{n,\text{CL}}$, the Donoho–Stahel estimators of location and scatter, denoted by $T_{n,\text{DS}}$ and the S -estimators of location and scatter, denoted by $T_{n,\text{S}}$. Both robust estimators are calibrated to attain 50% breakdown point. The corresponding frequencies of rejection are reported in Tables 1 and 2, where $\Delta = 0$ corresponds to the observations generated according to the null hypothesis.

Taking as reference the first row of Table 1, as expected, we observe some loss of power of the classical test based on $T_{n,\text{CL}}$ under the alternatives of the distributions considered in $H_0^{(2)}$ and $H_0^{(4)}$, where the data follow heavier tailed distributions. On the other hand, the opposite is observed when considering $\mathcal{U}(\mathcal{S}_p)$ and $\mathcal{U}(\mathcal{B}_p)$. The extreme situation is found under $H_0^{(7)}$ and its alternatives, since the classical test completely loses its power. Indeed, in this case this test is unable to distinguish between data coming from the elliptical distribution $\mathcal{T}_{2,1}(\mathbf{0}, \mathbf{I})$ from data

generated under its alternatives $H_{1,\Delta}^{(7)}$ for $\Delta = 0.5, 1$ and 1.5 . Besides, Table 2 shows that for both families of robust estimators of location and scatter matrix similar results are obtained, either in level or power. Indeed, with both robust estimators, the proposed tests lead to comparable results to those obtained with the classical test for $H_0^{(j)}$ for $j = 1, 3, 5$ and 6 , even when for $j = 1$ and 3 there is some loss of power under $H_{1,0.5}^{(j)}$. However, the robust tests outperform the behaviour observed with $T_{n,CL}$ under the alternatives of $H_0^{(2)}$ and $H_0^{(4)}$, getting larger frequencies of rejection. Finally, from Table 2 we conclude that using $T_{n,DS}$ and $T_{n,S}$ the decision rule has a good performance and is informative even under $H_0^{(7)}$ and its alternative hypotheses.

We have also considered two other alternatives also studied in Batsidis *et al.* (2014) which gave power 1 as $H_1^{\star(1)}$ and $H_1^{\star(2)}$, for that reason the results are omitted in the Tables. One of the alternatives, denoted as $H_1^{\star(3)}$, is obtained generating random vectors with two independent components with a common beta distribution $Be(5, 1)$. The other one, $H_1^{\star(4)}$, corresponds to the distribution of a random vector with distribution $0.5 N_p(\mathbf{0}, \mathbf{I}) + 0.5 N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (1, 2)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}.$$

Δ	$T_{n,CL}$					
	0	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$
		0.5	1	1.5		
$H_0^{(1)}$	0.060	0.478	1.000	1.000	1.000	1.000
$H_0^{(2)}$	0.086	0.220	0.748	0.888	1.000	1.000
$H_0^{(3)}$	0.048	0.412	0.994	0.998	1.000	1.000
$H_0^{(4)}$	0.044	0.110	0.794	0.988	1.000	1.000
$H_0^{(5)}$	0.057	0.892	1.000	1.000	1.000	1.000
$H_0^{(6)}$	0.054	1.000	1.000	1.000	1.000	1.000
$H_0^{(7)}$	0.056	0.058	0.062	0.076	1.000	1.000

Table 1: Frequency of rejection for the bootstrap test $T_{n,CL}$ for $n = 200$ and dimension $p = 2$. \star indicates that the frequency of rejection is significantly different from the nominal level.

As noted before, the exact sizes of the test statistics, i.e., $\pi_{H_0}(T_{n,CL})$, $\pi_{H_0}(T_{n,DS})$ and $\pi_{H_0}(T_{n,S})$ fluctuate around the fixed level $\alpha = 0.05$. To help in the visual comparison of the power performance of the three test statistics, as in Batsidis *et al.* (2014), Table 3 reports the size-corrected relative exact powers $\rho_{H_1}(T_{n,DS}, T_{n,CL})$ and $\rho_{H_1}(T_{n,S}, T_{n,CL})$. For two test statistics, $T_n^{(1)}$ and $T_n^{(2)}$, $\rho_{H_1}(T_n^{(1)}, T_n^{(2)})$ was defined in Morales *et al.* (2004) as

$$\rho_{H_1}(T_n^{(1)}, T_n^{(2)}) = \left(\frac{D_{H_1}(T_n^{(1)})}{D_{H_1}(T_n^{(2)})} - 1 \right) \times 100, \quad (7)$$

Δ	$T_{n,DS}$						$T_{n,S}$					
	0	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$	0	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$
		0.5	1	1.5				0.5	1	1.5		
$H_0^{(1)}$	0.048	0.256	0.990	1.000	1.000	1.000	0.060	0.270	0.988	1.000	1.000	1.000
$H_0^{(2)}$	0.058	0.283	0.984	1.000	1.000	1.000	0.060	0.309	0.980	1.000	1.000	1.000
$H_0^{(3)}$	0.042	0.244	0.984	1.000	1.000	1.000	0.050	0.264	0.986	1.000	1.000	1.000
$H_0^{(4)}$	0.056	0.206	0.862	0.998	1.000	1.000	0.056	0.212	0.876	0.998	1.000	1.000
$H_0^{(5)}$	0.062	0.608	1.000	1.000	1.000	1.000	0.068	0.552	1.000	1.000	1.000	1.000
$H_0^{(6)}$	0.048	1.000	1.000	1.000	1.000	1.000	0.050	1.000	1.000	1.000	1.000	1.000
$H_0^{(7)}$	0.048	0.066	0.464	0.832	1.000	1.000	0.046	0.068	0.440	0.840	1.000	1.000

Table 2: Frequency of rejection for the bootstrap test $T_{n,DS}$ and $T_{n,S}$ for $n = 200$ and dimension $p = 2$. \star indicates that the frequency of rejection is significantly different from the nominal level.

with $D_{H_1}(T_n) = \pi_{H_1}(T_n) - \pi_{H_0}(T_n)$. This measure allows to clarify the fluctuations in the powers which are more difficult to observe in Tables 1 and 2.

Δ	$\rho_{H_1}(T_{n,DS}, T_{n,CL})$						$\rho_{H_1}(T_{n,S}, T_{n,CL})$				
	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$		$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$
	0.5	1	1.5				0.5	1	1.5		
$H_0^{(1)}$	-50.239	0.213	1.277	1.277	1.277		-49.761	-1.277	0.000	0.000	0.000
$H_0^{(2)}$	67.910	39.879	17.456	3.064	3.064		85.821	38.973	17.207	2.845	2.845
$H_0^{(3)}$	-44.506	-0.423	0.842	0.630	0.630		-41.209	-1.057	0.000	-0.210	-0.210
$H_0^{(4)}$	127.273	7.467	-0.212	-1.255	-1.255		136.364	9.333	-0.212	-1.255	-1.255
$H_0^{(5)}$	-34.611	-0.530	-0.530	-0.530	-0.530		-42.036	-1.167	-1.167	-1.167	-1.167
$H_0^{(6)}$	0.634	0.634	0.634	0.634	0.634		0.423	0.423	0.423	0.423	0.423
$H_0^{(7)}$	800.000	6833.333	3820.000	0.848	0.848		1000.000	6466.667	3870.000	1.059	1.059

Table 3: Size corrected relative exact power for the robust bootstrap tests $T_{n,DS}$ and $T_{n,S}$ with respect to the classical one $T_{n,CL}$ for $n = 200$ and dimension $p = 2$.

Table 3 shows that in most cases, larger values of $\rho_{H_1}(T_n, T_{n,CL})$ are obtained with the Donoho–Stahel estimators over the S –estimators leading to the conclusion that the test based on the Donoho–Stahel estimators is a preferable choice. As noted before, the robust tests outperform the classical one specially for alternatives close to the null hypothesis under $H_0^{(2)}$, $H_0^{(4)}$, $H_0^{(6)}$ and $H_0^{(7)}$. This performance was expected for the heavy tailed distributions $H_0^{(2)}$ and $H_0^{(7)}$, but it is also present under moderate tails as those of the $\mathcal{T}_{p,3}(\mathbf{0}, \mathbf{I})$ since fourth moments do not exist. On the other hand, as expected, the classical test has a superior behaviour under a Gaussian distribution, when $\Delta = 0.5$. The better performance of $T_{n,CL}$ for $\Delta = 0.5$, is also observed under $H_0^{(3)}$ and the uniform distribution over the unit circle. With respect to the detection of the alternatives $H_1^{\star(1)}$ and $H_1^{\star(2)}$, all procedures are almost equivalent. Based on the simulated results obtained for the considered distributions, we recommend the test statistic based on the Donoho–Stahel estimators.

For this reason, in dimension $p = 5$ we only compare the test statistics $T_{n,DS}$ and $T_{n,CL}$.

6.2 Simulation study in dimension $p = 5$

In order to compare the performance of the two test statistics $T_{n,CL}$ and $T_{n,DS}$, under the null hypothesis, we generate n independent observations $\mathbf{Z}_1, \dots, \mathbf{Z}_n, \mathbf{Z}_i \sim P, \mathbf{Z}_i \in \mathbb{R}^5$, following different elliptical distributions as follows $H_0^{(1)} : P = N_p(\mathbf{0}, \mathbf{I})$, $H_0^{(2)} : P$ is the Pearson type II distribution generated as $\sqrt{V}\mathbf{U}$ where $\mathbf{U} \sim \mathcal{U}(\mathcal{S}_p)$ and $V \sim Be(p/2, m)$, with $m = 3/2$, $H_0^{(3)} : P = \mathcal{T}_{p,5}(\mathbf{0}, \mathbf{I})$ and $H_0^{(4)} : P = \mathcal{T}_{p,1}(\mathbf{0}, \mathbf{I})$. As in Section 6.1, we consider observations $\mathbf{X}_i, i = 1, \dots, n$ generated under the alternative hypotheses $H_{1,\Delta}^{(j)}$, with $\Delta = 0.25, 0.5, 0.75, 1$ and 1.5 . Besides, we studied the performance under four fixed alternatives $H_1^{\star(j)}$ for $j = 1$ to 4 defined as follows. Under $H_1^{\star(1)}$, the data have the distribution of a random vector with p independent components, the first $p - 1$ having distribution $\mathcal{E}(1)$ and the last one $N(0, 1)$. This distribution corresponds to $H_1^{\star(1)}$ in dimension $p = 2$. The second fixed alternative $H_1^{\star(2)}$ corresponds to the distribution of a random vector \mathbf{X} with p independent components each of them with distribution $\mathcal{E}(1)$. Under $H_1^{\star(3)}$, $\mathbf{X}_i \sim \mathbf{X}$ where \mathbf{X} is a random vector with p independent components with common distribution $Be(5, 1)$. Finally, $H_1^{\star(4)}$ corresponds to the situation in which \mathbf{X} has p independent components, the first $p - 1$ with common distribution $\mathcal{E}(1)$ and the last one $\mathcal{T}_{p,1}$.

The frequencies of rejection are reported in Table 4, where $\Delta = 0$ corresponds to the observations generated according to the null hypothesis. Besides, Table 5 reports the size-corrected relative exact powers $\rho_{H_1}(T_{n,DS}, T_{n,CL})$ as defined in (7).

Δ	$T_{n,CL}$									
	$H_{1,\Delta}^{(j)}$						$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.060	0.088	0.646	0.998	1.000	1.000	1.000	1.000	1.000	1.000
$H_0^{(2)}$	0.046	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_0^{(3)}$	0.080*	0.070	0.290	0.866	0.996	1.000	1.000	1.000	1.000	1.000
$H_0^{(4)}$	0.058	0.078	0.080	0.086	0.094	0.118	1.000	1.000	1.000	1.000
Δ	$T_{n,DS}$									
	$H_{1,\Delta}^{(j)}$						$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.034	0.044	0.428	0.984	1.000	1.000	1.000	1.000	1.000	1.000
$H_0^{(2)}$	0.036	0.982	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_0^{(3)}$	0.040	0.052	0.270	0.870	0.996	1.000	1.000	1.000	1.000	1.000
$H_0^{(4)}$	0.054	0.056	0.126	0.398	0.766	0.994	1.000	1.000	1.000	1.000

Table 4: Frequency of rejection for the bootstrap test $T_{n,CL}$ and $T_{n,DS}$ for $n = 200$ and dimension $p = 5$, $\alpha = 0.05$. \star indicates that the frequency of rejection is significantly different from the nominal level.

From Table 4, one observes that, except for the Pearson distribution, the observed level of the classical procedure, $\pi_{H_0}(T_{n,\text{CL}})$, is slightly higher than the nominal one. However, $T_{n,\text{CL}}$ leads to a liberal test only for the $\mathcal{T}_{p,5}(\mathbf{0}, \mathbf{I})$ distribution. On the contrary, $\pi_{H_0}(T_{n,\text{DS}})$ is smaller than the nominal level, except for the Cauchy distribution in which case, the exact size is close to 0.05. However, in none of the considered situations the observed frequencies of rejection are significantly different from the nominal level $\alpha = 0.05$. For the Cauchy distribution, as expected, the classical test is non-informative when considering the alternatives $H_{1,\Delta}^{(4)}$. On the other hand, both procedures detect the alternatives $H_1^{\star(j)}$ for $j = 1, \dots, 4$. Table 5 shows the advantage of the procedure based on $T_{n,\text{DS}}$ over that based on $T_{n,\text{CL}}$, except for $H_{1,0.05}^{(1)}$. The inadequate behavior of $T_{n,\text{CL}}$ for the Cauchy distribution shown by a power almost equal to the level for the alternatives $H_{1,\Delta}^{(4)}$ is more clear when comparing the values of the size-corrected relative exact powers $\rho_{H_1}(T_{n,\text{DS}}, T_{n,\text{CL}})$.

Δ	$\rho_{H_1}(T_{n,\text{DS}}, T_{n,\text{CL}})$								
	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	-64.286	-32.765	1.279	2.766	2.766	2.766	2.766	2.766	2.766
$H_0^{(2)}$	-0.421	1.048	1.048	1.048	1.048	1.048	1.048	1.048	1.048
$H_0^{(3)}$	-220.000	9.524	5.598	4.367	4.348	4.348	4.348	4.348	4.348
$H_0^{(4)}$	-90.000	227.273	1128.571	1877.778	1466.667	0.425	0.425	0.425	0.425

Table 5: Size corrected relative exact power of the robust bootstrap test $T_{n,\text{DS}}$ with respect to the classical $T_{n,\text{CL}}$ one, when $n = 200$, $p = 5$ and $\alpha = 0.05$.

6.3 Comparisons with other tests for elliptical symmetry

Taking into account the better performance of the procedure based on $T_{n,\text{DS}}$ over that based on $T_{n,\text{CL}}$ and $T_{n,\text{S}}$, in this section, we compare the conditional test based on $T_{n,\text{DS}}$ with some other methods found in the literature. The simulation conditions are similar to those described in Sections 6.1 and 6.2.

As mentioned in Section 2, there is a wide literature on methods to test for elliptical symmetry. According to the simulation power studies performed in Huffer and Park (2007), none of the tests introduced in Manzotti *et al.* (2002), Schott (2002) or Huffer and Park (2007) is uniformly superior for detecting departures from the null hypothesis. On the other hand, Batsidis *et al.* (2014) also showed that their proposal is comparable in power to that defined in Schott (2002).

The purpose of the numerical study in this section, is to show that the proposed test statistic is a useful option to the previously defined methods, in particular when moments do not exist.

Since there is no superior test statistic, we decided to choose for the comparison a test statistic which can be easily computed and has a tractable null distribution. For this reason, we excluded the test defined in Koltchinskii and Sakhanenko (2000) as well as the statistic defined in Beran (1979). With respect to the test proposed in Batsidis and Zografos (2013), their method helps to decide departures from a specific elliptical model, while our procedure is designed to detect departures from the whole family of elliptical distributions. For this reason, it is not included in the comparison.

On the contrary, the test defined by Schott (2002) is easy to compute since it is based on a fourth moment statistic denoted as $T_{n,\text{SCH}}$. Besides, this statistic is asymptotically χ_ν^2 , where ν depends on the dimension of the data, but not on the underlying null elliptical distribution. Based on the simulation studies reported in Schott (2002), Huffer and Park (2007) and Batsidis *et al.* (2014), the test based on $T_{n,\text{SCH}}$ has observed level close to the nominal one and good empirical power. Besides, as our procedure, the test statistic is affine invariant. It is worth noting, that the asymptotic behavior of $T_{n,\text{SCH}}$ is derived for distributions having finite moments up to order eight so that it will be sensitive to departures from this assumption, even if the distribution is elliptical.

We also include in the comparison the test statistic, $T_{n,\text{BAT}}$, recently introduced in Batsidis *et al.* (2014) that is based on a power divergence family of statistics depending on a parameter λ . According to the simulation results in Batsidis *et al.* (2014), we select $\lambda = 1$ (similar results were obtained for $\lambda = 2/3$). As noted by these authors, $T_{n,\text{BAT}}$ has approximately a chi-square distribution with degrees of freedom depending on the sample size and may be computed in a simple way. As mentioned in Batsidis *et al.* (2014), this test statistic has a very good power for a variety of alternatives, even when it is not affine invariant.

The observed frequency of rejection and the size corrected relative exact powers for $T_{n,\text{BAT}}$ are given in Tables 6 and 7 for $p = 2$ and in Tables 8 and 9, when $p = 5$. Analogous quantities for $T_{n,\text{SCH}}$ are reported in Tables 10 and 11 for $p = 2$ and in Tables 12 and 13 for $p = 5$.

As expected, when the underlying distribution has no moments, the test based on $T_{n,\text{SCH}}$ becomes non informative, since it relies on the assumption of existence of eight order moments. The same happens when the data are generated according to the multivariate Student distributions $\mathcal{T}_{2,3}(\mathbf{0}, \mathbf{I})$ and $\mathcal{T}_{5,5}(\mathbf{0}, \mathbf{I})$ since these distributions do not satisfy the assumptions in Schott (2002). When $p = 2$, in the four other situations, even if the test reaches in most cases the desired level, it has difficulties to detect the selected alternatives, specially the alternatives $H_1^{\star(3)}$ and $H_1^{\star(4)}$. In dimension $p = 5$,

Δ	$T_{n,BAT}$							
	0	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
		0.5	1	1.5				
$H_0^{(1)}$	0.016★	0.048	0.512	0.980	0.456	1.000	0.974	0.976
$H_0^{(2)}$	0.570★	0.550	0.740	0.970	0.456	1.000	0.974	0.976
$H_0^{(3)}$	0.054	0.088	0.522	0.976	0.456	1.000	0.974	0.976
$H_0^{(4)}$	0.140★	0.122	0.418	0.852	0.456	1.000	0.974	0.976
$H_0^{(5)}$	0.014★	0.032	0.842	0.998	0.456	1.000	0.974	0.976
$H_0^{(6)}$	0.030	0.832	1.000	1.000	0.456	1.000	0.974	0.976
$H_0^{(7)}$	0.926★	0.892	0.904	0.892	0.456	1.000	0.974	0.976

Table 6: Frequency of rejection for the test defined in Batsidis *et al.* (2014) for $n = 200$ and dimension $p = 2$ with $\lambda = 1$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

Δ	$\rho_{H_1}(T_{n,BAT}, T_{n,DS})$						
	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.5	1	1.5				
$H_0^{(1)}$	-84.615	-47.346	1.261	-53.782	3.361	0.630	0.840
$H_0^{(2)}$	-108.889	-81.642	-57.537	-112.102	-54.352	-57.113	-56.900
$H_0^{(3)}$	-83.168	-50.319	-3.758	-58.038	-1.253	-3.967	-3.758
$H_0^{(4)}$	-112.000	-65.509	-24.416	-66.525	-8.898	-11.653	-11.441
$H_0^{(5)}$	-96.703	-11.727	4.904	-52.879	5.117	2.345	2.559
$H_0^{(6)}$	-15.756	1.891	1.891	-55.252	1.891	-0.840	-0.630
$H_0^{(7)}$	-288.889	-105.289	-104.337	-149.370	-92.227	-94.958	-94.748

Table 7: Size corrected relative exact power for test based on the statistic $T_{n,BAT}$ defined in Batsidis *et al.* (2014) with respect to the robust bootstrap test $T_{n,DS}$ for $n = 200$ and dimension $p = 2$ with $\lambda = 1$.

Δ	$T_{n,BAT}$								
	0	$H_{1,\Delta}^{(j)}$				$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
		0.25	0.5	0.75	1	1.5			
$H_0^{(1)}$	0.026	0.030	0.060	0.186	0.630	0.998	0.936	1.000	0.888
$H_0^{(2)}$	0.020★	0.080	0.970	1.000	1.000	1.000	0.936	1.000	0.888
$H_0^{(3)}$	0.036	0.036	0.050	0.210	0.516	0.972	0.936	1.000	0.888
$H_0^{(4)}$	0.800★	0.804	0.798	0.808	0.810	0.846	0.936	1.000	0.888

Table 8: Frequency of rejection for the test defined in Batsidis *et al.* (2014) for $n = 200$ and dimension $p = 5$ with $\lambda = 1$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

for the Pearson distribution, the level and power performance of $T_{n,SCH}$ is analogous to that of our procedure, while for the normal distribution, the test based on $T_{n,DS}$ has a better detection power except for $\Delta = 0.25$. For the chosen fixed alternatives, $H_1^{\star(j)}$, $j = 3, 4$ our procedure has much better power.

With respect to the proposal given in Batsidis *et al.* (2014), the test based on $T_{n,BAT}$ becomes non-informative when contaminating with a Cauchy distribution or when the data follow a Cauchy distribution, which can be explained by the fact that the test is based on the sample mean and the

Δ	$\rho_{H_1}(T_{n,BAT}, T_{n,DS})$								
	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	-60.000	-91.371	-83.158	-37.474	0.621	-5.797	0.828	-10.766	-0.207
$H_0^{(2)}$	-93.658	-1.452	1.660	1.660	1.660	-4.979	1.660	-9.959	0.622
$H_0^{(3)}$	-100.000	-93.913	-79.036	-49.791	-2.500	-6.250	0.417	-11.250	-0.625
$H_0^{(4)}$	100.000	-102.778	-97.674	-98.596	-95.106	-85.624	-78.858	-90.698	-79.915

Table 9: Size corrected relative exact power for test based on the statistic $T_{n,BAT}$ defined in Batsidis *et al.* (2014) with respect to the robust bootstrap test $T_{n,DS}$ for $n = 200$ and dimension $p = 5$ with $\lambda = 1$.

Δ	$T_{n,SCH}$							
	0	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
		0.5	1	1.5				
$H_0^{(1)}$	0.032	0.074	0.218	0.312	0.548	0.282	0.140	0.864
$H_0^{(2)}$	0.504★	0.490	0.444	0.394	0.548	0.282	0.140	0.864
$H_0^{(3)}$	0.054	0.050	0.182	0.296	0.548	0.282	0.140	0.864
$H_0^{(4)}$	0.098★	0.094	0.094	0.158	0.548	0.282	0.140	0.864
$H_0^{(5)}$	0.064	0.112	0.310	0.380	0.548	0.282	0.140	0.864
$H_0^{(6)}$	0.054	0.276	0.382	0.410	0.548	0.282	0.140	0.864
$H_0^{(7)}$	0.396★	0.394	0.410	0.432	0.548	0.282	0.140	0.864

Table 10: Frequency of rejection for the test defined in Schott (2002) for $n = 200$ and dimension $p = 2$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

	$\rho_{H_1}(T_{n,\text{SCH}}, T_{n,\text{DS}})$						
	$H_{1,\Delta}^{(j)}$			$H_1^{(1)}$	$H_1^{*(2)}$	$H_1^{*(3)}$	$H_1^{*(4)}$
Δ	0.5	1	1.5				
$H_0^{(1)}$	-79.808	-80.255	-70.588	-45.798	-73.740	-88.656	-12.605
$H_0^{(2)}$	-106.222	-106.480	-111.677	-95.329	-123.567	-138.641	-61.783
$H_0^{(3)}$	-101.980	-86.412	-74.739	-48.434	-76.200	-91.023	-15.449
$H_0^{(4)}$	-102.667	-100.496	-93.631	-52.331	-80.509	-95.551	-18.856
$H_0^{(5)}$	-91.209	-73.774	-66.311	-48.401	-76.759	-91.898	-14.712
$H_0^{(6)}$	-76.681	-65.546	-62.605	-48.109	-76.050	-90.966	-14.916
$H_0^{(7)}$	-111.111	-96.635	-95.408	-84.034	-111.975	-126.891	-50.840

Table 11: Size corrected relative exact power for test based on the statistic $T_{n,SCH}$ defined in Schott (2002) with respect to the robust bootstrap test $T_{n,DS}$ for $n = 200$ and dimension $p = 2$.

sample covariance matrix. On the other hand, for the number of replications considered the level 0.05 is not attained for the normal distribution and for data uniformly distributed over the unit circle, in dimension $p = 2$. This fact was also observed in Table 10 of Batsidis *et al.* (2014) for the normal distribution with 10000 replications. Table 7 shows that, as for the Schott's test, our method outperforms the procedure based on $T_{n,BAT}$ in the majority of the cases considered. Similar conclusions for the considered distributions and alternatives are obtained in dimension $p = 5$, as shown in Table 8.

	$T_{n,\text{SCH}}$									
		$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
Δ	0	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.040	0.064	0.182	0.550	0.790	0.900	0.864	0.872	0.314	0.516
$H_0^{(2)}$	0.056	0.468	0.900	0.944	0.954	0.964	0.864	0.872	0.314	0.516
$H_0^{(3)}$	0.054	0.046	0.052	0.150	0.410	0.800	0.864	0.872	0.314	0.516
$H_0^{(4)}$	0.974★	0.974	0.974	0.974	0.970	0.946	0.864	0.872	0.314	0.516

Table 12: Frequency of rejection for the test defined in Schott (2002) for $n = 200$ and dimension $p = 5$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

Δ	$\rho_{H_1}(T_{n,\text{SCH}}, T_{n,\text{DS}})$									
	0.25	0.5	$H_{1,\Delta}^{(j)}$			$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$	
			0.75	1	1.5					
$H_0^{(1)}$	140.000	-63.959	-46.316	-22.360	-10.973	-14.700	-13.872	-71.636	-50.725	
$H_0^{(2)}$	-56.448	-12.448	-7.884	-6.847	-5.809	-16.183	-15.353	-73.237	-52.282	
$H_0^{(3)}$	-166.667	-100.870	-88.434	-62.762	-22.292	-15.625	-14.792	-72.917	-51.875	
$H_0^{(4)}$	-100.000	-100.000	-100.000	-100.562	-102.979	-111.628	-110.782	-169.767	-148.414	

Table 13: Size corrected relative exact power for test based on the statistic $T_{n,\text{SCH}}$ defined in Schott (2002) with respect to the robust bootstrap test $T_{n,\text{DS}}$ for $n = 200$ and dimension $p = 5$.

These facts are highlighted in Tables 11 and 13 that report the size corrected relative powers $\rho_{H_1}(T_{n,\text{SCH}}, T_{n,\text{DS}})$. The negative values reported in all cells confirm the better performance of $T_{n,\text{DS}}$. Note that even if, for the Cauchy distribution, the test proposed in Batsidis *et al.* (2014) has a positive value when $\Delta = 0.25$ and $p = 5$, the test is non-informative having power almost constant for $H_{1,\Delta}^{(4)}$. Hence, in this case, the size corrected size does not provide a good measure to compare the test statistics.

It is worth noticing that our conclusions regarding the better performance of the test based on the Donoho–Stahel estimators are valid only for the considered distributions and alternatives. A more extensive simulation study would be necessary to conclude that, in general, $T_{n,\text{DS}}$ should be preferred. This interesting comparison may be object of future work.

6.4 Simulation study in dimension $p = 5$ with $n = 50$.

In Section 5, we noticed that for small sample sizes the distribution of the test statistic does not allow to distinguish the elliptical distributions from those considered in the alternative. For that reason, in the simulation study reported in Sections 6.1 and 6.2, we choose as sample size $n = 200$. To complement the results obtained in Section 6.2 and to study the effect of a smaller sample size on the decisions taken, we report here the observed frequencies of rejection for the test based on

the sample mean and covariance matrix, $T_{n,CL}$ and for that based on the Donoho–Stahel estimators, $T_{n,DS}$, when $p = 5$ and $n = 50$, which represents a challenging situation due to the ratio between sample size and dimension. As in Section 6.3, we also compare their performance with that of the test statistic, $T_{n,BAT}$, introduced in Batsidis *et al.* (2014) and with the test defined by Schott (2002), $T_{n,SCH}$. The simulation conditions as well as the considered alternatives are described in Section 6.2.

The corresponding frequencies of rejection are reported in Tables 14, 15 and 16, where $\Delta = 0$ corresponds to the observations generated according to the null hypothesis. Besides, Table 17 reports the size-corrected relative exact powers $\rho_{H_1}(T_{n,DS}, T_{n,CL})$ as defined in (7), while Tables 18 and 19 report the size corrected relative powers $\rho_{H_1}(T_{n,BAT}, T_{n,DS})$ and $\rho_{H_1}(T_{n,SCH}, T_{n,DS})$. Note that a positive value of size corrected relative power $\rho_{H_1}(T_{n,1}, T_{n,2})$ indicates that the test based on $T_{n,1}$ has a better detection capability than that based on $T_{n,2}$ and the size of its advantage is quantified by $\rho_{H_1}(T_{n,1}, T_{n,2})$. Similarly, a negative value of $\rho_{H_1}(T_{n,1}, T_{n,2})$ provides a measure of the deficiency of $T_{n,1}$ with respect to $T_{n,2}$.

	$T_{n,CL}$									
	$H_{1,\Delta}^{(j)}$						$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
Δ	0	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.054	0.068	0.180	0.528	0.844	0.988	1.000	1.000	1.000	0.998
$H_0^{(2)}$	0.044	0.440	0.984	1.000	1.000	1.000	1.000	1.000	1.000	0.998
$H_0^{(3)}$	0.128★	0.134	0.186	0.370	0.652	0.940	1.000	1.000	1.000	0.998
$H_0^{(4)}$	0.228★	0.230	0.240	0.276	0.290	0.362	1.000	1.000	1.000	0.998
	$T_{n,DS}$									
	$H_{1,\Delta}^{(j)}$						$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
Δ	0	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.052	0.050	0.086	0.236	0.542	0.898	0.992	1.000	0.972	0.980
$H_0^{(2)}$	0.060	0.204	0.858	0.992	1.000	1.000	0.992	1.000	0.972	0.980
$H_0^{(3)}$	0.068	0.050	0.090	0.190	0.360	0.774	0.992	1.000	0.972	0.980
$H_0^{(4)}$	0.044	0.036	0.048	0.098	0.150	0.330	0.992	1.000	0.972	0.980

Table 14: Frequency of rejection for the bootstrap test $T_{n,CL}$ and $T_{n,DS}$ for $n = 50$ and dimension $p = 5$, $\alpha = 0.05$. \star indicates that the frequency of rejection is significantly different from the nominal level.

Table 14 shows that, even for this small sample size, the robust procedures allow to detect the considered alternatives keeping the exact sizes of the test statistic, i.e., $\pi_{H_0}(T_{n,DS})$ around the nominal level $\alpha = 0.05$. As in Section 6.2 and 6.3, we indicate with a \star those cases, in which the observed empirical frequencies of rejection are different from the nominal level with a significance level $\gamma = 0.01$. As expected, the test based on the sample mean and covariance matrix outperforms that based on the Donoho–Stahel estimators under the normal distribution due to the

Δ	$T_{n,BAT}$									
	0	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
		0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.016★	0.016	0.030	0.040	0.084	0.248	0.152	0.302	0.144	0.274
$H_0^{(2)}$	0.016★	0.036	0.186	0.440	0.544	0.640	0.152	0.302	0.144	0.274
$H_0^{(3)}$	0.016★	0.024	0.036	0.036	0.074	0.184	0.152	0.302	0.144	0.274
$H_0^{(4)}$	0.438★	0.408	0.362	0.374	0.352	0.324	0.152	0.302	0.144	0.274

Table 15: Frequency of rejection for the test defined in Batsidis *et al.* (2014) for $n = 50$ and dimension $p = 5$ with $\lambda = 1$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

Δ	$T_{n,SCH}$									
	0	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
		0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	0.048	0.034	0.034	0.076	0.142	0.312	0.238	0.258	0.074	0.708
$H_0^{(2)}$	0.060	0.074	0.270	0.432	0.514	0.586	0.238	0.258	0.074	0.708
$H_0^{(3)}$	0.038	0.036	0.044	0.060	0.084	0.182	0.238	0.258	0.074	0.708
$H_0^{(4)}$	0.964★	0.966	0.966	0.956	0.948	0.900	0.238	0.258	0.074	0.708

Table 16: Frequency of rejection for the test defined in Schott (2002) for $n = 50$ and dimension $p = 5$. ★ indicates that the frequency of rejection is significantly different from the nominal level.

loss of efficiency of the robust estimators. The advantage of $T_{n,CL}$ is also observed for the Pearson distribution, in particular, when $\Delta = 0.25$ and $\Delta = 0.5$. These two facts are consistent with the behaviour described in Section 5, where for dimension $p = 2$ and $n = 50$, the distribution of the test statistic has troubles to distinguish between the null hypothesis and close alternatives for most of the considered elliptical distributions. Note that the test based on the sample mean and covariance matrix becomes liberal under $\mathcal{T}_{p,5}(\mathbf{0}, \mathbf{I})$. Besides, the test based on the robust estimators shows its advantage for this distribution, except for $\Delta = 0.25$ where $T_{n,DS}$ does not succeed in detecting the hypothesis, leading to a large negative value on the size corrected relative power. On the other hand, under the Cauchy distribution $T_{n,CL}$ is non-informative, while $T_{n,DS}$ is able to distinguish all the alternatives except when $\Delta = 0.25$ and 0.5 . These facts become more evident in Table 17, where most size corrected relative powers are positive for distributions different from the normal. The large negative value obtained at $\Delta = 0.25$ for the Cauchy distribution can be explained by means of two facts. The first one is that $T_{n,CL}$ has power almost constant, so that the denominator is close to 0, while the second one is that the power of $T_{n,DS}$ decreases at $\Delta = 0.25$ with respect to its size. Note that, given two test statistics $T_{n,1}$ and $T_{n,2}$, when the test based on $T_{n,2}$ is non-informative, a negative value of the size corrected relative power $\rho_{H_1}(T_{n,1}, T_{n,2})$ does not provide a good measure to conclude the benefits of $T_{n,2}$ over $T_{n,1}$.

With respect to the test statistics, $T_{n,\text{BAT}}$ and $T_{n,\text{SCH}}$, both procedures loose their capability of detection under the Cauchy distribution, since their behaviour relies on the existence of moments. On the other hand, the test statistic proposed in Batsidis *et al.* (2014) is conservative in all situations except for the Cauchy distribution ($H_0^{(4)}$), where it is liberal. The same conclusions obtained when $n = 200$ are preserved in the actual setting, that is, the procedure proposed in this paper outperforms these competitors in the majority of the situations considered. As mentioned in Section 6.3, our conclusions on the benefits of $T_{n,\text{DS}}$ are valid only for the considered distributions and alternatives. Quite surprisingly, even for this small sample size the procedure based on $T_{n,\text{DS}}$ shows a reasonable performance probably due to the bootstrap method used to compute the p -value.

Δ	$\rho_{H_1}(T_{n,\text{DS}}, T_{n,\text{CL}})$								
	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	-114.286	-73.016	-61.181	-37.975	-9.422	-0.634	0.211	-2.748	-1.695
$H_0^{(2)}$	-63.636	-15.106	-2.510	-1.674	-1.674	-2.510	-1.674	-4.603	-3.564
$H_0^{(3)}$	-400.000	-62.069	-49.587	-44.275	-13.054	5.963	6.881	3.670	4.828
$H_0^{(4)}$	-500.000	-66.667	12.500	70.968	113.433	22.798	23.834	20.207	21.548

Table 17: Size corrected relative exact power of the robust bootstrap test $T_{n,\text{DS}}$ with respect to the classical $T_{n,\text{CL}}$ one, when $n = 50$, $p = 5$ and $\alpha = 0.05$.

Δ	$\rho_{H_1}(T_{n,\text{BAT}}, T_{n,\text{DS}})$								
	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	-100.000	-58.824	-86.957	-86.122	-72.577	-85.532	-69.831	-86.087	-72.198
$H_0^{(2)}$	-86.111	-78.697	-54.506	-43.830	-33.617	-85.408	-69.574	-85.965	-71.957
$H_0^{(3)}$	-144.444	-9.091	-83.607	-80.137	-76.204	-85.281	-69.313	-85.841	-71.711
$H_0^{(4)}$	275.000	-2000.000	-218.519	-181.132	-139.860	-130.169	-114.226	-131.681	-117.949

Table 18: Size corrected relative exact power for test based on the statistic $T_{n,\text{BAT}}$ defined in Batsidis *et al.* (2014) with respect to the robust bootstrap test $T_{n,\text{DS}}$ for $n = 50$ and dimension $p = 5$ with $\lambda = 1$.

Δ	$\rho_{H_1}(T_{n,\text{SCH}}, T_{n,\text{DS}})$								
	$H_{1,\Delta}^{(j)}$					$H_1^{\star(1)}$	$H_1^{\star(2)}$	$H_1^{\star(3)}$	$H_1^{\star(4)}$
	0.25	0.5	0.75	1	1.5				
$H_0^{(1)}$	600.000	-141.176	-84.783	-80.816	-68.794	-79.787	-77.848	-97.174	-28.879
$H_0^{(2)}$	-90.278	-73.684	-60.086	-51.702	-44.043	-80.901	-78.936	-98.465	-29.565
$H_0^{(3)}$	-88.889	-72.727	-81.967	-84.247	-79.603	-78.355	-76.395	-96.018	-26.535
$H_0^{(4)}$	-125.000	-50.000	-114.815	-115.094	-122.378	-176.582	-173.849	-195.905	-127.350

Table 19: Size corrected relative exact power for test based on the statistic $T_{n,\text{SCH}}$ defined in Schott (2002) with respect to the robust bootstrap test $T_{n,\text{DS}}$ for $n = 50$ and dimension $p = 5$.

Acknowledgement This research was partially supported by Grants W276 and 20120130100241BA from

Universidad of Buenos Aires, PIP 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT, Argentina and also received financial support from Portuguese National Funds through FCT (Fundação para a Ciência e a Tecnologia) under the scope of project PEst-OE/MAT/UI0822/2011. The authors wish to thank two anonymous referees for valuable comments which led to an improved version of the original paper.

Appendix A: Numerical computation of the test statistic

The defined test statistic $T_{n,\mathbf{m},\mathbf{V}}$ involves an integral that may be calculated numerically. In dimension 2, the approximation described below is easy to perform. Assume that w has compact support $\mathcal{I} = [-b, b]$ and split it in a grid of $N_{\mathcal{I}}$ points t_i . We consider M random directions $\{\mathbf{a}_j\}_{j=1}^M$ in \mathcal{S}_p generated according to a uniform distribution on the sphere. Once the robust estimates \mathbf{m}_n and \mathbf{V}_n are obtained from the sample, for each t_i on the grid and each generated random direction \mathbf{a}_j , we compute $I_{ij} = \left\{ \sqrt{n} P_n \sin \left[t_i \mathbf{a}_j^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] \right\}^2 w(t_i)$. Then, we approximate the desired test statistic by $2b \text{sur}(\mathcal{S}_p) \sum_{i=1}^{N_{\mathcal{I}}} \sum_{j=1}^M I_{ij} / (N_{\mathcal{I}} M)$, where $\text{sur}(\mathcal{S}_p)$ denotes the surface area of the sphere in \mathbb{R}^p of radius 1.

To get an alternative expression for the test statistic, we will restrict our attention to the situation where $w(t) = \mathbb{I}_{[-b,b]}(t)/(2b)$. Let $Z_i(\mathbf{a}) = \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{m}_n)$, then

$$\int (\sqrt{n} P_n \{\sin(t Z_i(\mathbf{a}))\})^2 w(t) dt = \frac{1}{2b n} \sum_{i,j} \int_{-b}^b \sin(t Z_i(\mathbf{a})) \sin(t Z_j(\mathbf{a})) dt.$$

Using that $\sin(x) \sin(y) = (\cos(x-y) - \cos(x+y))/2$ and denoting $U_{ij}^+(\mathbf{a}) = Z_i(\mathbf{a}) + Z_j(\mathbf{a}) = \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i + \mathbf{X}_j - 2\mathbf{m}_n)$ and $U_{ij}^-(\mathbf{a}) = Z_i(\mathbf{a}) - Z_j(\mathbf{a}) = \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{X}_j)$, we get that

$$\begin{aligned} \int (\sqrt{n} P_n \{\sin(t Z_i(\mathbf{a}))\})^2 w(t) dt &= \frac{1}{2b} \frac{1}{2n} \sum_{i,j} \int_{-b}^b \cos(t U_{ij}^-(\mathbf{a})) dt - \int_{-b}^b \cos(t U_{ij}^+(\mathbf{a})) dt \\ &= \frac{1}{2b n} \sum_{i,j} \frac{\sin(b U_{ij}^-(\mathbf{a}))}{U_{ij}^-(\mathbf{a})} - \frac{\sin(b U_{ij}^+(\mathbf{a}))}{U_{ij}^+(\mathbf{a})}, \end{aligned}$$

which implies that

$$T_{n,\mathbf{m},\mathbf{V}} = \frac{1}{2n} \sum_{i,j} \left[\mathbb{E}_{\mathcal{U}_p} \frac{\sin(b \mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{X}_j))}{b \mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{X}_j)} - \mathbb{E}_{\mathcal{U}_p} \frac{\sin(b \mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i + \mathbf{X}_j - 2\mathbf{m}_n))}{b \mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i + \mathbf{X}_j - 2\mathbf{m}_n)} \right],$$

where $\mathbf{U} = (U_1, \dots, U_p)^T \sim \mathcal{U}(\mathcal{S}_p) = \mathcal{U}_p$. Define $(D_{ij}^+)^2 = d^2(\mathbf{X}_i - \boldsymbol{\mu}, -(\mathbf{X}_j - \boldsymbol{\mu}), \mathbf{V})$ and $(D_{ij}^-)^2 = d^2(\mathbf{X}_i - \boldsymbol{\mu}, \mathbf{X}_j - \boldsymbol{\mu}, \mathbf{V})$, where $d^2(\mathbf{x}, \mathbf{v}, \boldsymbol{\Sigma}) = (\mathbf{x} - \mathbf{v})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{v})$ is the squared Mahalanobis distance.

Since $\mathbf{U} \sim \mathcal{U}(\mathcal{S}_p)$, we have that $Y = (p-1)^{\frac{1}{2}} U_1 / \sqrt{1 - U_1^2} \sim \mathcal{T}_{1, p-1}$ (see Muirhead, 1982, pp.38) and $U_1 = Y / \sqrt{p-1 + Y^2}$. Moreover, since $\mathbf{U}^T \mathbf{v} \sim U_1$ for any $\mathbf{v} \in \mathcal{S}_p$ we have that $\mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i - \mathbf{X}_j) \sim U_1 D_{ij}^-$ and $\mathbf{U}^T \mathbf{V}_n^{-1/2} (\mathbf{X}_i + \mathbf{X}_j - 2\mathbf{m}_n) \sim U_1 D_{ij}^+$. These facts entail that

$$T_{n, \mathbf{m}, \mathbf{V}} = \frac{1}{2n} \sum_{i,j} \left[f(b D_{ij}^-) - f(b D_{ij}^+) \right],$$

where the function $f : \mathbb{R} \rightarrow \mathbb{R}$ is defined as

$$f(u) = \mathbb{E}_{\mathcal{T}_{1, p-1}} \left(\frac{\sin \left(u \frac{Y}{\sqrt{p-1+Y^2}} \right)}{\frac{Y}{\sqrt{p-1+Y^2}} u} \right) = \mathbb{E}_{\mathcal{U}_p} \left(\frac{\sin(u U_1)}{u U_1} \right),$$

with $\sin(u)/u = 1$ if $u = 0$. Then, using that $D_{ij}^- = D_{ji}^-$, $D_{ij}^+ = D_{ji}^+$, $D_{ii}^- = 0$ and $f(0) = 1$, we get a simpler expression for the test statistic given by

$$T_{n, \mathbf{m}, \mathbf{V}} = \frac{1}{2n} \sum_{j=1}^n \left\{ 1 - f(b D_{jj}^+) + 2 \sum_{i=1}^{j-1} \left[f(b D_{ij}^-) - f(b D_{ij}^+) \right] \right\}.$$

To summarize, in order to compute the test statistic, the user only has to evaluate the function f by Monte Carlo over a grid of points.

Appendix B: Proofs of Theorems 3.1 and 3.2

PROOF OF THEOREM 3.1. Using that $\sin(x-y) = \sin(x) \cos(y) - \cos(x) \sin(y)$ and since $\mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) = \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) - \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu})$, we have that $\sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] = \sqrt{n} W_{1,n}(t, \mathbf{a}) - \sqrt{n} W_{2,n}(t, \mathbf{a})$ where

$$\begin{aligned} W_{1,n}(t, \mathbf{a}) &= \cos \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu}) \right] P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] \\ W_{2,n}(t, \mathbf{a}) &= \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu}) \right] P_n \cos \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right]. \end{aligned}$$

Denote as $Z_n(t, \mathbf{a}, \mathbf{A}) = P_n \cos [t \mathbf{a}^T \mathbf{A} (\mathbf{X} - \boldsymbol{\mu})]$ and $\zeta(t, \mathbf{a}, \mathbf{A}) = \mathbb{E}_P \cos [t \mathbf{a}^T \mathbf{A} (\mathbf{X} - \boldsymbol{\mu})]$. Note that since $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, we have that $\zeta(t, \mathbf{a}, \mathbf{A}) = \psi(t^2 \mathbf{a}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T \mathbf{a})$ which entails that $\zeta(t, \mathbf{a}, \boldsymbol{\Sigma}^{-1/2}) =$

$\psi(t^2)$. The Dominated Convergence Theorem implies that

$$\lim_{\mathbf{A} \rightarrow \Sigma^{-1/2}} \mathbb{E}_P \sup_{t \in \mathbb{R}, \mathbf{a} \in \mathcal{S}_p} \left| \cos \left[t \mathbf{a}^T \mathbf{A} (\mathbf{X} - \boldsymbol{\mu}) \right] - \cos \left(t \mathbf{a}^T \Sigma^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right) \right| = 0,$$

which, together with the fact that $\mathbf{V}_n \xrightarrow{p} \Sigma$, entails that $\sup_{t \in \mathbb{R}, \mathbf{a} \in \mathcal{S}_p} |\zeta(t, \mathbf{a}, \mathbf{V}_n^{-1/2}) - \zeta(t, \mathbf{a}, \Sigma^{-1/2})| \xrightarrow{p} 0$, that is,

$$\sup_{t \in \mathbb{R}, \mathbf{a} \in \mathcal{S}_p} \left| \mathbb{E}_P \cos \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] - \psi(t^2) \right| \xrightarrow{p} 0. \quad (\text{B.1})$$

Let $\|\cdot\|_s$ be a norm in the space \mathcal{M} of symmetric positive definite matrices. Then, as $\mathbf{V}_n \xrightarrow{p} \Sigma$, we have that for n large enough with high probability, $\|\mathbf{V}_n^{-1/2} - \Sigma^{-1/2}\|_s \leq 1$. Let $\mathcal{F} = \{f(\mathbf{x}) = \cos(t \mathbf{a}^T \mathbf{A}(\mathbf{x} - \boldsymbol{\mu}))\}$, $t \in \mathcal{I}$, $\mathbf{a} \in \mathcal{S}_p$, $\mathbf{A} \in \mathbb{R}^{p \times p} : \|\mathbf{A} - \Sigma^{-1/2}\|_s \leq 1\}$ and $\mathcal{G}_C = \{f_{\mathbf{b}}(\mathbf{x}) = \cos(\mathbf{b}^T(\mathbf{x} - \boldsymbol{\mu}))\}$, $\mathbf{b} \in \mathbb{R}^p : \|\mathbf{b}\| \leq C\}$. Then, for some $C > 0$, $\mathcal{F} \subset \mathcal{G}_C$. Note that the functions $f_{\mathbf{b}}(\mathbf{x})$ are such that the map $\mathbf{b} \mapsto f_{\mathbf{b}}(\mathbf{x})$ is continuous for all fixed \mathbf{x} and its envelope $F(\mathbf{x}) = \sup_{\|\mathbf{b}\| \leq C} |f_{\mathbf{b}}(\mathbf{x})|$ satisfies that $F \in L_1(P)$ since $F \leq 1$. Hence, using Lemma 3.10 in van de Geer (2000), we have that the class \mathcal{G}_C has finite bracketing number $N_{[]}(\epsilon, \mathcal{G}_C, L_1(P))$, which entails that \mathcal{G}_C is a Glivenko–Cantelli class of functions, that is, $\sup_{\|\mathbf{b}\| \leq C} |P_n \cos[\mathbf{b}^T(\mathbf{X} - \boldsymbol{\mu})] - P \cos[\mathbf{b}^T(\mathbf{X} - \boldsymbol{\mu})]| \xrightarrow{a.s.} 0$. This convergence implies that $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |Z_n(t, \mathbf{a}, \mathbf{V}_n^{-1/2}) - \zeta(t, \mathbf{a}, \mathbf{V}_n^{-1/2})| \xrightarrow{p} 0$. Hence, using (B.1), we get that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |Z_n(t, \mathbf{a}, \mathbf{V}_n^{-1/2}) - \psi(t^2)| \xrightarrow{p} 0. \quad (\text{B.2})$$

From the fact that $\lim_{u \rightarrow 0} \sin u / u = 1$, the consistency of \mathbf{V}_n and \mathbf{m}_n , we get that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \frac{\sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu}) \right]}{t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu})} - 1 \right| \xrightarrow{p} 0. \quad (\text{B.3})$$

Using that $\sqrt{n}(\mathbf{m}_n - \boldsymbol{\mu}) = O_{\mathbb{P}}(1)$, together with (B.2) and (B.3), we conclude that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \sqrt{n} W_{2,n}(t, \mathbf{a}) - \psi(t^2) \left[t \mathbf{a}^T \Sigma^{-1/2} \sqrt{n}(\mathbf{m}_n - \boldsymbol{\mu}) \right] \right| \xrightarrow{p} 0,$$

which together with the fact that \mathbf{m}_n admits a Bahadur expansion leads to

$$\sqrt{n} W_{2,n}(t, \mathbf{a}) = \psi(t^2) t \mathbf{a}^T \Sigma^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) + \alpha_{\mathbf{m}} \left(\|\Sigma^{-1/2} (\mathbf{X}_i - \boldsymbol{\mu})\| \right) + R_{2,n}(t, \mathbf{a}), \quad (\text{B.4})$$

where $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |R_{2,n}(t, \mathbf{a})| \xrightarrow{p} 0$.

Similarly, using that $|\cos(u) - 1| \leq |u|$, that \mathcal{I} is a bounded interval and the consistency of \mathbf{V}_n and \mathbf{m}_n , we obtain that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \cos \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{m}_n - \boldsymbol{\mu}) \right] - 1 \right| \xrightarrow{p} 0. \quad (\text{B.5})$$

Hence, the proof will be complete if we show that

- a) $\sqrt{n} W_{3,n}(t, \mathbf{a}) = \sqrt{n} P_n \sin \left[t \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right]$ converges to a Gaussian process and
- b) $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] - \sqrt{n} P_n \sin \left[t \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] \right| \xrightarrow{p} 0.$

Effectively, if a) and b) hold the process $\sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right]$ is tight, so, using (B.5), we can write $\sqrt{n} W_{1,n}(t, \mathbf{a}) = \sqrt{n} W_{3,n}(t, \mathbf{a}) + R_{1,n}(t, \mathbf{a})$, where $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |R_{1,n}(t, \mathbf{a})| \xrightarrow{p} 0$, which together with (B.4) leads to

$$\begin{aligned} \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_n^{-1/2} (\mathbf{X} - \mathbf{m}_n) \right] &= \psi(t^2) t \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \alpha_{\mathbf{m}} \left(\|\boldsymbol{\Sigma}^{-1/2} (\mathbf{X}_i - \boldsymbol{\mu})\| \right) \\ &\quad + \sqrt{n} P_n \sin \left[t \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] + R_n(t, \mathbf{a}), \end{aligned}$$

where $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |R_n(t, \mathbf{a})| \xrightarrow{p} 0$.

The proof of a) follows from Ghosh and Ruymgaart (1992), so it only remains to show b).

To derive b), denote $Z_n^*(t, \mathbf{a}, \mathbf{A}) = P_n \sin [t \mathbf{a}^T \mathbf{A} (\mathbf{X} - \boldsymbol{\mu})]$. Then, we have that $\mathbb{E}_P Z_n^*(t, \mathbf{a}, \mathbf{A}) = 0$, since $\mathbf{X} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$. Note that $\mathcal{F}^* = \{f(\mathbf{x}) = \sin(t \mathbf{a}^T \mathbf{A} (\mathbf{x} - \boldsymbol{\mu})), t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p, \mathbf{A} \in \mathbb{R}^{p \times p} : \|\mathbf{A} - \boldsymbol{\Sigma}^{-1/2}\|_s \leq 1\} \subset \mathcal{G}_C^*$ for some $C > 0$ where $\mathcal{G}_C^* = \{f_{\mathbf{b}}(\mathbf{x}) = \sin(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{x} - \boldsymbol{\mu})), \mathbf{b} \in \mathbb{R}^p : \|\mathbf{b}\| \leq C\}$. In the Appendix C, it is shown that \mathcal{G}_C^* is Donsker, which entails the uniform equicontinuity leading to b). \square

PROOF THEOREM 3.2. As in Theorem 3.1, let $\mathcal{W}_n(t, \mathbf{a}) = \sqrt{n} P_n \sin \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{X} - \mathbf{m}_{n, \mathbf{X}}) \right]$. We will show that the process $\mathcal{W}_n = \{\mathcal{W}_n(t, \mathbf{a}), (t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p\}$ converges in distribution to the Gaussian process $\mathcal{W}^* = \{\mathcal{W}^*(t, \mathbf{a}), (t, \mathbf{a}) \in \mathcal{I} \times \mathcal{S}_p\}$ with $\mathcal{W}^*(t, \mathbf{a}) = \mathcal{W}(t, \mathbf{a}) + (1/\ell!) B_\ell(t, \mathbf{a})$.

As in the proof of Theorem 3.1, we have that $\mathcal{W}_n(t, \mathbf{a}) = \sqrt{n} W_{1,n}(t, \mathbf{a}) - \sqrt{n} W_{2,n}(t, \mathbf{a})$ where

$$\begin{aligned} W_{1,n}(t, \mathbf{a}) &= \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) \right] P_n \sin \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] \\ W_{2,n}(t, \mathbf{a}) &= \sin \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) \right] P_n \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right]. \end{aligned}$$

Besides, we also have that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) \right] - 1 \right| \xrightarrow{p} 0, \quad (\text{B.6})$$

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| \frac{\sin \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) \right]}{t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu})} - 1 \right| \xrightarrow{p} 0, \quad (\text{B.7})$$

hold since $\mathbf{V}_{n, \mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$ and $\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu} \xrightarrow{p} 0$. On the other hand, from (6) and using that $\mathbf{m}_{n, \mathbf{Z}}$ satisfies (3) together with the fact that $\mathbf{V}_{n, \mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$, we get that

$$\sqrt{n} \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\Sigma}^{-1/2} (\mathbf{Z}_i - \boldsymbol{\mu}) \alpha_{\mathbf{m}} \left(\|\boldsymbol{\Sigma}^{-1/2} (\mathbf{Z}_i - \boldsymbol{\mu})\| \right) + R_n, \quad (\text{B.8})$$

where $R_n \xrightarrow{p} 0$. Hence, if we show that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} \left| P_n \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] - \psi(t^2) \right| \xrightarrow{p} 0, \quad (\text{B.9})$$

we obtain the following expansion for $\sqrt{n} W_{2, n}(t, \mathbf{a})$

$$\sqrt{n} W_{2, n}(t, \mathbf{a}) = \psi(t^2) t \mathbf{a}^T \boldsymbol{\Sigma}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{Z}_i - \boldsymbol{\mu}) \alpha_{\mathbf{m}} \left(\|\boldsymbol{\Sigma}^{-1/2} (\mathbf{Z}_i - \boldsymbol{\mu})\| \right) + R_{2, n}(t, \mathbf{a}),$$

where $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |R_{2, n}(t, \mathbf{a})| \xrightarrow{p} 0$.

To obtain (B.9), notice that from the proof of Theorem 3.1, we have that

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |P_n \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{Z} - \boldsymbol{\mu}) \right] - \psi(t^2)| \xrightarrow{p} 0, \quad (\text{B.10})$$

since $\mathbf{Z} \sim \mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ and $\mathbf{V}_{n, \mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$. Besides, using that $|\cos(u) - \cos(v)| \leq |u - v|$ and $X_i = \mathbf{Z}_i + \mathbf{Y}_i/n^\alpha$, we get the bound

$$\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |P_n \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] - P_n \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{Z} - \boldsymbol{\mu}) \right]| \leq \lambda_{n, \max}^{1/2} \frac{\nu}{n^\alpha} P_n \|\mathbf{Y}\|,$$

where $\lambda_{n, \max}$ stands for the largest eigenvalue of the matrix $\mathbf{V}_{n, \mathbf{X}}^{-1}$. Therefore, using that $\mathbb{E}\|\mathbf{Y}\| < \infty$ and $\mathbf{V}_{n, \mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$, we get (B.9) from (B.10).

Denote as $M_{n, \mathbf{X}}(t, \mathbf{a}, \mathbf{A}) = P_n \sin[t \mathbf{a}^T \mathbf{A} (\mathbf{X} - \boldsymbol{\mu})]$ and $M_{n, \mathbf{Z}}(t, \mathbf{a}, \mathbf{A}) = P_n \sin[t \mathbf{a}^T \mathbf{A} (\mathbf{Z} - \boldsymbol{\mu})]$.

Then, noting that $W_{1,n}(t, \mathbf{a}) = \cos \left[t \mathbf{a}^T \mathbf{V}_{n, \mathbf{X}}^{-1/2} (\mathbf{m}_{n, \mathbf{X}} - \boldsymbol{\mu}) \right] M_{n, \mathbf{X}}(t, \mathbf{a}, \mathbf{V}_{n, \mathbf{X}}^{-1/2})$, from (B.6), we obtain that $\sqrt{n} W_{1,n}(t, \mathbf{a}) = M_{n, \mathbf{X}}(t, \mathbf{a}, \mathbf{V}_{n, \mathbf{X}}^{-1/2}) + R_{1,n}(t, \mathbf{a})$, with $\sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |R_{1,n}(t, \mathbf{a})| \xrightarrow{p} 0$, if

$$\sqrt{n} \sup_{t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p} |M_{n, \mathbf{X}}(t, \mathbf{a}, \mathbf{V}_{n, \mathbf{X}}^{-1/2})| = O_{\mathbb{P}}(1) \quad (\text{B.11})$$

holds.

Recall that, from the proof of Theorem 3.1, $\sqrt{n} M_{n, \mathbf{Z}}(t, \mathbf{a}, \mathbf{V}_{n, \mathbf{X}}^{-1/2})$ converges to a Gaussian process since $\mathbf{V}_{n, \mathbf{X}} \xrightarrow{p} \boldsymbol{\Sigma}$. Let δ be such that $\|\mathbf{A} - \boldsymbol{\Sigma}^{-1/2}\|_s \leq \delta$ entails that $|\lambda_{\max}(\mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T) - 1| \leq \delta_0 / \nu$ where $\delta_0 = \min_{1 \leq m \leq \ell} (\delta_m)$ and δ_m are given in assumption f). Denote $\mathcal{F} = \{f(\mathbf{y}, \mathbf{z}) = (t \mathbf{a}^T \mathbf{A} \mathbf{y})^\ell \sin^{(\ell)}(t \mathbf{a}^T \mathbf{A}(\mathbf{z} - \boldsymbol{\mu}))\}$, $((t, \mathbf{a}, \mathbf{A}) \in \mathcal{A})$, where $\mathcal{A} = \{(t, \mathbf{a}, \mathbf{A}) : t \in \mathcal{I}, \mathbf{a} \in \mathcal{S}_p, \|\mathbf{A} - \boldsymbol{\Sigma}^{-1/2}\|_s \leq \delta\}$. Then, the proof will be completed if we show the following convergences

$$\sup_{(t, \mathbf{a}, \mathbf{A}) \in \mathcal{A}} \left| \sqrt{n} \{M_{n, \mathbf{X}}(t, \mathbf{a}, \mathbf{A}) - M_{n, \mathbf{Z}}(t, \mathbf{a}, \mathbf{A})\} - \frac{1}{\ell!} P_n \left\{ (t \mathbf{a}^T \mathbf{A} \mathbf{Y})^\ell \sin^{(\ell)}(t \mathbf{a}^T \mathbf{A}(\mathbf{Z} - \boldsymbol{\mu})) \right\} \right| \xrightarrow{p} 0, \quad (\text{B.12})$$

$$\sup_{f \in \mathcal{F}} |P_n f(\mathbf{Y}, \mathbf{Z}) - P f(\mathbf{Y}, \mathbf{Z})| \xrightarrow{p} 0, \quad (\text{B.13})$$

for some δ small enough, since (B.12) and (B.13) entail (B.11) and also the desired expansion. For any $1 \leq s \leq \ell$, define

$$\mathcal{G}_s = \{g_{\mathbf{b}}(\mathbf{y}, \mathbf{z}) = (\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{y})^s \sin^{(s)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})), \mathbf{b} \in \mathbb{R}^p : \|\mathbf{b}\| \leq \nu + \delta_0\}. \quad (\text{B.14})$$

The proof of (B.13) follows using Lemma 3.10 of van de Geer (2000) and similar arguments to those considered in proof of Theorem 3.1 applied to the classes of functions \mathcal{F} and \mathcal{G}_ℓ since $\mathcal{F} \subset \mathcal{G}_\ell$ and the envelope $G(\mathbf{y}, \mathbf{z}) = \sup_{\|\mathbf{b}\| \leq C} |g_{\mathbf{b}}(\mathbf{y}, \mathbf{z})| \leq C^\ell \|\mathbf{y}\|^\ell \in L^1(P)$ with $C = (\nu + \delta_0) \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}^{-1})$.

It only remains to prove (B.12) which follow if we show that $\sup_{\|\mathbf{b}\| \leq \nu + \delta_0} |Y_n(\mathbf{b})| \xrightarrow{p} 0$ where

$$\begin{aligned} Y_n(\mathbf{b}) &= \sqrt{n} \left\{ P_n \sin \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X} - \boldsymbol{\mu}) \right] - P_n \sin \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{Z} - \boldsymbol{\mu}) \right] \right\} \\ &\quad - \frac{1}{\ell!} P_n \left\{ \left(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y} \right)^\ell \sin^{(\ell)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})) \right\}. \end{aligned}$$

Using a Taylor's expansion and noting that $\mathbf{X}_i - \boldsymbol{\mu} = (\mathbf{Z}_i - \boldsymbol{\mu}) + \mathbf{Y}_i / n^\alpha$ and $\alpha = 1/(2\ell)$ we have that $(1/n) \sum_{i=1}^n \sin \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{X}_i - \boldsymbol{\mu}) \right] = (1/n) \sum_{i=1}^n \sin \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} (\mathbf{Z}_i - \boldsymbol{\mu}) \right] + S_{1,n} + S_{2,n} + S_{3,n}$,

where

$$\begin{aligned}
S_{1,n} &= \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^{\ell-1} \frac{1}{s!} \frac{1}{n^{\alpha s}} \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \right]^s \sin^{(s)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}_i - \boldsymbol{\mu})) = \sum_{s=1}^{\ell-1} \frac{1}{s!} \frac{1}{n^{\alpha s}} S_{1,n,s} \\
S_{2,n} &= \frac{1}{n^{\alpha \ell}} \frac{1}{\ell!} \frac{1}{n} \sum_{i=1}^n \left(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \right)^\ell \sin^{(\ell)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}_i - \boldsymbol{\mu})) \\
S_{3,n} &= \frac{1}{n^{\alpha(\ell+1)}} \frac{1}{(\ell+1)!} \frac{1}{n} \sum_{i=1}^n \left(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \right)^{\ell+1} \sin^{(\ell+1)}(\xi_n)
\end{aligned}$$

with $\xi_n = \theta_n \mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}_i - \boldsymbol{\mu}) + (1 - \theta_n) \mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})$, for some $\theta_n \in (0, 1)$. Thus, $Y_n(\mathbf{b}) = \sqrt{n}(S_{1,n} + S_{3,n})$ so to conclude the proof, we only have to show that $\sqrt{n} \sup_{\|\mathbf{b}\| \leq \delta_0 + \nu} |S_{j,n}| \xrightarrow{p} 0$, for $j = 1, 3$. Using that $\ell\alpha = 1/2$, we get that, for any $\|\mathbf{b}\| \leq \delta_0 + \nu$,

$$|\sqrt{n} S_{3,n}| \leq \frac{1}{n^{\alpha}} \frac{1}{(\ell+1)!} P_n \left| \left(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y} \right)^{\ell+1} \right| \leq \frac{1}{n^{\alpha}} \frac{1}{(\ell+1)!} C^{\ell+1} P_n \|\mathbf{Y}\|^{\ell+1}$$

where $C = (\nu + \delta_0) \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}^{-1})$, which entails that $\sqrt{n} S_{3,n} \xrightarrow{p} 0$ since $\mathbb{E} \|\mathbf{Y}\|^{\ell+1} < \infty$.

To obtain that $\sqrt{n} \sup_{\|\mathbf{b}\| \leq C} |S_{1,n}| \xrightarrow{p} 0$, it is enough to show that, for any $1 \leq s \leq \ell - 1$, $\sqrt{n} \sup_{\|\mathbf{b}\| \leq C} |S_{1,n,s}| / n^{\alpha s} \xrightarrow{p} 0$ which will follow if we prove that

$$\sqrt{n} \sup_{\|\mathbf{b}\| \leq C} |S_{1,n,s}| = O_{\mathbb{P}}(1). \quad (\text{B.15})$$

Note that by assumption f), for any $s < \ell$, $\mathbb{E} \left\{ \left[\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}_i \right]^s \sin^{(s)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}_i - \boldsymbol{\mu})) \right\} = 0$, when $\|\mathbf{b}\| \leq \delta_0 + \nu$. Thus, $\mathbb{E} S_{1,n,s} = 0$ holds, so (B.15) follows from the fact that the class \mathcal{G}_s defined in (B.14) is Donsker which is derived in the Appendix C. \square

Appendix C: Proof that $\mathcal{G}_C^* = \{f_{\mathbf{b}}(\mathbf{x}) = \sin(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}))\}$, $\mathbf{b} \in \mathbb{R}^p$: $\|\mathbf{b}\| \leq C\}$ and \mathcal{G}_s defined in (B.14) are Donsker

When $\mathbb{E}_P \|\mathbf{X} - \boldsymbol{\mu}\|^2 < \infty$, the result follows easily from Lemma 2.5 in van de Geer (2000) and Theorem 2.7.11 in van der Vaart and Wellner (1996). However, since we do not assume this moment condition, we have to work more carefully and we will use the fact that $\int_0^1 \sqrt{\log H_u} du < \infty$.

To provide a unified proof, denote as

$$\mathcal{G}_s = \{g_{\mathbf{b}}(\mathbf{y}, \mathbf{z}) = (\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2} \mathbf{y})^s \sin^{(s)}(\mathbf{b}^T \boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})), \mathbf{b} \in \mathbb{R}^p : \|\mathbf{b}\| \leq C\}.$$

Hence, when $s = 0$, $\mathcal{G}_s = \mathcal{G}_C^*$ while for $C = \nu + \delta_0$ we get the class defined in (B.14). It is then, enough to show that \mathcal{G}_s is Donsker when $\int_0^1 \sqrt{\log H_u} du < \infty$ if $s = 0$ or if \mathbf{Y} and \mathbf{Z} are independent while, if $1 \leq s < \ell$ and \mathbf{Y} and \mathbf{Z} are not independent we will use that for any $1 \leq s < \ell$, $\int_0^1 \sqrt{\log H_{u^q}} du < \infty$, where $q = \ell/(\ell - s)$, $\sin^{(s)}(u) = (-1)^s \sin(u)$ or $\sin^{(s)}(u) = (-1)^{s+1} \cos(u)$ and that $\mathbb{E}\|\mathbf{Y}\|^{2s} < \infty$.

For simplicity, denote $\mathcal{B}_p(\mathbf{b}, \delta) = \{\mathbf{u} : \|\mathbf{u} - \mathbf{b}\| \leq \delta\}$, $\mathcal{B}_p(\delta) = \mathcal{B}_p(\mathbf{0}, \delta)$, $\|f\|_{L^2(P)} = (\mathbb{E}_P f^2)^{1/2}$, $\lambda = \sqrt{\lambda_{\max}(\boldsymbol{\Sigma}^{-1})}$ and $A = 2\lambda^s \left\{ 3 (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} + C^s (\mathbb{E}\|\mathbf{Y}\|^{2\ell})^{\frac{s}{2\ell}} \right\}$.

For any fixed $\rho > 0$ and $\boldsymbol{\beta} \in \mathcal{B}_p(C)$, define $\Psi(\mathbf{y}, \mathbf{z}, \boldsymbol{\beta}, \rho) = \sup_{\{\mathbf{b} \in \mathcal{B}_p(C) \cap \mathcal{B}_p(\boldsymbol{\beta}, \rho)\}} |g_{\mathbf{b}}(\mathbf{y}, \mathbf{z}) - g_{\boldsymbol{\beta}}(\mathbf{y}, \mathbf{z})|$. Note that the continuity of the sinus entails that the supremum can be taken over \mathbb{Q}^p , so that $\Psi(\mathbf{x}, \boldsymbol{\beta}, \rho)$ is measurable for each $\boldsymbol{\beta}$ and $\rho > 0$. Note that $|\sin^{(s)}(u) - \sin^{(s)}(v)| \leq |u - v|$ entails that $|g_{\mathbf{b}_1}(\mathbf{y}, \mathbf{z}) - g_{\mathbf{b}_2}(\mathbf{y}, \mathbf{z})| \leq \lambda^s \|\mathbf{y}\|^s \max(\|\mathbf{b}_1 - \mathbf{b}_2\|, \|\mathbf{b}_1 - \mathbf{b}_2\|^s) \left\{ 1 + C^s \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{z} - \boldsymbol{\mu})\| \right\}$. Then, if $S = \|\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\|$ we obtain the bound

$$\mathbb{E}_P \Psi^2(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \rho) \mathbb{I}_{[0, M]}(S) \leq \mathbb{E}\|\mathbf{Y}\|^{2s} \lambda^{2s} \{1 + C^s M\}^2 \max(\rho^2, \rho^{2s}). \quad (\text{C.1})$$

On the other hand, since $|g_{\mathbf{b}_1}(\mathbf{y}, \mathbf{z}) - g_{\mathbf{b}_2}(\mathbf{y}, \mathbf{z})| \leq \lambda^s \|\mathbf{y}\|^s \{\|\mathbf{b}_1 - \mathbf{b}_2\|^s + 2C^s\}$, we have that $\Psi^2(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \rho) \leq \lambda^{2s} \{\rho^{2s} + 2C^s\}^2 \|\mathbf{Y}\|^{2s}$ and so, using the dominated convergence Theorem we get that $\mathbb{E}_P \Psi^2(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \rho) \rightarrow 0$ as $\rho \rightarrow 0$.

For a given $0 < \epsilon < \min(1, A)$, let $\eta = \epsilon/A$, choose M_ϵ as the smallest value such that $\mathbb{P}(S > M_\epsilon) = \mathbb{P}(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\| > M_\epsilon) \leq (\epsilon/A)^{2\ell/(\ell-s)}$ and define $\rho_\epsilon = \min\{\eta, \eta/[C^s M_\epsilon]\} < 1$. Then, since $\max(\rho_\epsilon^2, \rho_\epsilon^{2s}) = \rho_\epsilon^2$, we get that

$$\begin{aligned} \mathbb{E}_P \Psi^2(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}, \rho_\epsilon) \mathbb{I}_{[0, M_\epsilon]}(S) &\leq \mathbb{E}\|\mathbf{Y}\|^{2s} \lambda^{2s} \left\{ \min\left(\eta, \frac{\eta}{C^s M_\epsilon}\right) + C^s M_\epsilon \min\left(\eta, \frac{\eta}{C^s M_\epsilon}\right) \right\}^2 \\ &\leq \mathbb{E}\|\mathbf{Y}\|^{2s} \lambda^{2s} 4\eta^2 \end{aligned} \quad (\text{C.2})$$

Let $N = N(\rho_\epsilon, \mathcal{B}_p(C))$ the minimum number of balls of radius ρ_ϵ and center in $\mathcal{B}_p(C)$, needed to cover the set $\mathcal{B}_p(C)$. Then, N is at most twice the number of balls of radius ρ_ϵ needed to cover the set $\mathcal{B}_p(C)$ for which a bound is given in Lemma 2.5 in van de Geer (2000). Hence, there exist $\mathbf{b}_1, \dots, \mathbf{b}_N$,

$\mathbf{b}_j \in \mathcal{B}_p(C)$ such that $\mathcal{B}_p(C) \subset \cup_{j=1}^N \mathcal{B}_p(\mathbf{b}_j, \rho_\epsilon)$. Define $u_j(\mathbf{y}, \mathbf{z}) = g_{\mathbf{b}_j}(\mathbf{y}, \mathbf{z}) + \Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon)$ and $\ell_j(\mathbf{y}, \mathbf{z}) = g_{\mathbf{b}_j}(\mathbf{y}, \mathbf{z}) - \Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon)$. Then, for any $\mathbf{b} \in \mathcal{B}_p(C)$, there exists $1 \leq j \leq N$ such that $\mathbf{b} \in \mathcal{B}_p(\mathbf{b}_j, \rho_\epsilon)$, so that $|g_{\mathbf{b}}(\mathbf{y}, \mathbf{z}) - g_{\mathbf{b}_j}(\mathbf{y}, \mathbf{z})| \leq \Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon)$ which entails that $\ell_j(\mathbf{y}, \mathbf{z}) \leq g_{\mathbf{b}}(\mathbf{y}, \mathbf{z}) \leq u_j(\mathbf{y}, \mathbf{z})$. On the other hand, $\Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon) = \Psi_{1,j}(\mathbf{y}, \mathbf{z}) + \Psi_{2,j}(\mathbf{y}, \mathbf{z})$ where $\Psi_{1,j}(\mathbf{y}, \mathbf{z}) = \Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon) \mathbb{I}_{[0, M_\epsilon]}(\|\Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|)$ and $\Psi_{2,j}(\mathbf{y}, \mathbf{z}) = \Psi(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon) \mathbb{I}_{(M_\epsilon, +\infty)}(\|\Sigma^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|)$. Note that using (C.2), we get that

$$\|\Psi_{1,j}\|_{L^2(P)} \leq 2 (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s \eta = 2 (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s \frac{\epsilon}{A}.$$

On the other hand, the fact that $\Psi^2(\mathbf{y}, \mathbf{z}, \mathbf{b}_j, \rho_\epsilon) \leq \lambda^{2s} \{\rho_\epsilon^s + 2C^s\}^2 \|\mathbf{Y}\|^{2s}$ entails that

$$\begin{aligned} \|\Psi_{2,j}\|_{L^2(P)} &\leq \lambda^s \{\rho_\epsilon^s + 2C^s\} (\mathbb{E}\|\mathbf{Y}\|^{2s} \mathbb{I}_{S > M_\epsilon})^{1/2} \\ &\leq (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s \eta + 2\lambda^s C^s \{\mathbb{E}(\|\mathbf{Y}\|^{2s} \mathbb{I}_{S > M_\epsilon})\}^{1/2} \end{aligned} \quad (\text{C.3})$$

We will apply Hölder inequality with $p = \ell/s$, so that $1/p = s/\ell$ and $q = \ell/(\ell - s)$. Then

$$\mathbb{E}\|\mathbf{Y}\|^{2s} \mathbb{I}_{S > M_\epsilon} \leq (\mathbb{E}\|\mathbf{Y}\|^{2sp})^{1/p} (\mathbb{E} \mathbb{I}_{S > M_\epsilon})^{1/q} = \left\{ \mathbb{E}\|\mathbf{Y}\|^{2\ell} \right\}^{\frac{s}{\ell}} \{\mathbb{P}(S > M_\epsilon)\}^{\frac{(\ell-s)}{\ell}}$$

so, using that $\eta = \epsilon/A$ we get

$$\begin{aligned} \|\Psi_{2,j}\|_{L^2(P)} &\leq (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s \frac{\epsilon}{A} + 2\lambda^s C^s \left\{ \mathbb{E}\|\mathbf{Y}\|^{2\ell} \right\}^{\frac{s}{2\ell}} \{\mathbb{P}(S > M_\epsilon)\}^{\frac{(\ell-s)}{2\ell}} \\ &\leq \left\{ (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s + 2\lambda^s C^s \left\{ \mathbb{E}\|\mathbf{Y}\|^{2\ell} \right\}^{\frac{s}{2\ell}} \right\} \frac{\epsilon}{A} \end{aligned}$$

so that

$$\begin{aligned} \|u_j - \ell_j\|_{L^2(P)} &= 2\|\Psi_{1,j} + \Psi_{2,j}\|_{L^2(P)} \leq 2\|\Psi_{1,j}\|_{L^2(P)} + 2\|\Psi_{2,j}\|_{L^2(P)} \\ &\leq \left\{ 6 (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s + 2\lambda^s C^s \left\{ \mathbb{E}\|\mathbf{Y}\|^{2\ell} \right\}^{\frac{s}{2\ell}} \right\} \frac{\epsilon}{A} = \epsilon. \end{aligned}$$

Summarizing, we have shown that the bracketing number $N_{[]}(\epsilon, \mathcal{G}_s, L^2(P))$ is smaller or equal than $N(\rho_\epsilon, \mathcal{B}_p(C))$ which, from Lemma 2.5 in van de Geer (2000), may be bounded as

$$N(\rho_\epsilon, \mathcal{B}_p(C)) \leq 2 \left(\frac{4C + \rho_\epsilon}{\rho_\epsilon} \right)^p.$$

Note that if $\rho_\epsilon > C$, $N(\rho_\epsilon, \mathcal{B}_p(C)) \leq 2 \times 5^p$, otherwise $N(\rho_\epsilon, \mathcal{B}_p(C)) \leq 2 (5C/\rho_\epsilon)^p$. Thus,

$$N_{[]}(\epsilon, \mathcal{G}_s, L^2(P)) \leq 2 \max \left(5^p, \left[\frac{5C}{\rho_\epsilon} \right]^p \right) \leq 2 \max \left(5^p, \left[\frac{5C}{\epsilon} \right]^p, \left[5 C^{s+1} A M_\epsilon \frac{1}{\epsilon} \right]^p \right). \quad (\text{C.4})$$

Note that in (C.4) we can always assume that $M_\epsilon \geq 1$, otherwise we take $M_\epsilon = 1$ which gives an upper bound, so that $M_\epsilon = H_{(\epsilon/A)^{\ell/(\ell-s)}}$. Thus, if we denote as $A = p \log(5) + p \log(5C) + p \log(5 C^{s+1} A) + \log(2)$, we have the following bound

$$\begin{aligned} \int_0^1 \sqrt{\log(N_{[]}(\epsilon, \mathcal{G}_s, L^2(P)))} du &\leq \int_0^1 \sqrt{A + p \log M_u + 2p \log \left(\frac{1}{u} \right)} du \\ &\leq \sqrt{A} + \sqrt{p} \int_0^1 \sqrt{\log M_u} du + \sqrt{2p} \int_0^1 \sqrt{\log \left(\frac{1}{u} \right)} du < \infty, \end{aligned}$$

since by hypothesis $\int_0^1 \sqrt{\log M_u} du = \int_0^1 \sqrt{\log H_{(u/A)^q}} du = A \int_0^A \sqrt{\log H_{u^q}} du < \infty$, where $q = \ell/(\ell - s)$, concluding the proof.

Note that if $s = 0$, the condition $\int_0^1 \sqrt{\log H_u} du < \infty$ suffices to prove that \mathcal{G}_s is Donsker. Furthermore, if \mathbf{Y}_i y \mathbf{Z}_i are independent the assumption $\int_0^1 \sqrt{\log H_{u^q}} du < \infty$ is also weakened to $\int_0^1 \sqrt{\log H_u} du < \infty$. Indeed, in this case, we define $A = 2\lambda^s (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \{3 + C^s\}$ and we choose M_ϵ as the smallest value greater or equal than 1 such that $\mathbb{P}(S > M_\epsilon) = \mathbb{P}(\|\Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})\| > M_\epsilon) \leq \epsilon^2/A^2$, that is, $M_\epsilon = H_{(\epsilon/A)}$ and using the independence in (C.3), we get

$$\begin{aligned} \|\Psi_{2,j}\|_{L^2(P)} &\leq (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s \frac{\epsilon}{A} + 2\lambda^s C^s (\mathbb{E}\|\mathbf{Y}\|^{2s} \mathbb{P}(S > M_\epsilon))^{1/2} \\ &\leq \left\{ (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \lambda^s + 2\lambda^s C^s (\mathbb{E}\|\mathbf{Y}\|^{2s})^{1/2} \right\} \frac{\epsilon}{A}, \end{aligned}$$

so that $\|u_j - \ell_j\|_{L^2(P)} \leq \epsilon$. So, as above, the bracketing number $N_{[]}(\epsilon, \mathcal{G}_s, L^2(P))$ is smaller or equal than $N(\rho_\epsilon, \mathcal{B}_p(C))$ getting the bound (C.4). Hence, if we denote as $A = p \log(5) + p \log(5C) + p \log(5 C^{s+1} A) + \log(2)$, we obtain

$$\int_0^1 \sqrt{\log(N_{[]}(\epsilon, \mathcal{G}_s, L^2(P)))} du \leq \sqrt{A} + \sqrt{p} \int_0^1 \sqrt{\log M_u} du + \sqrt{2p} \int_0^1 \sqrt{\log \left(\frac{1}{u} \right)} du < \infty$$

since by hypothesis $\int_0^1 \sqrt{\log M_u} du = \int_0^1 \sqrt{\log H_{(u/A)}} du = A \int_0^A \sqrt{\log H_u} du < \infty$.

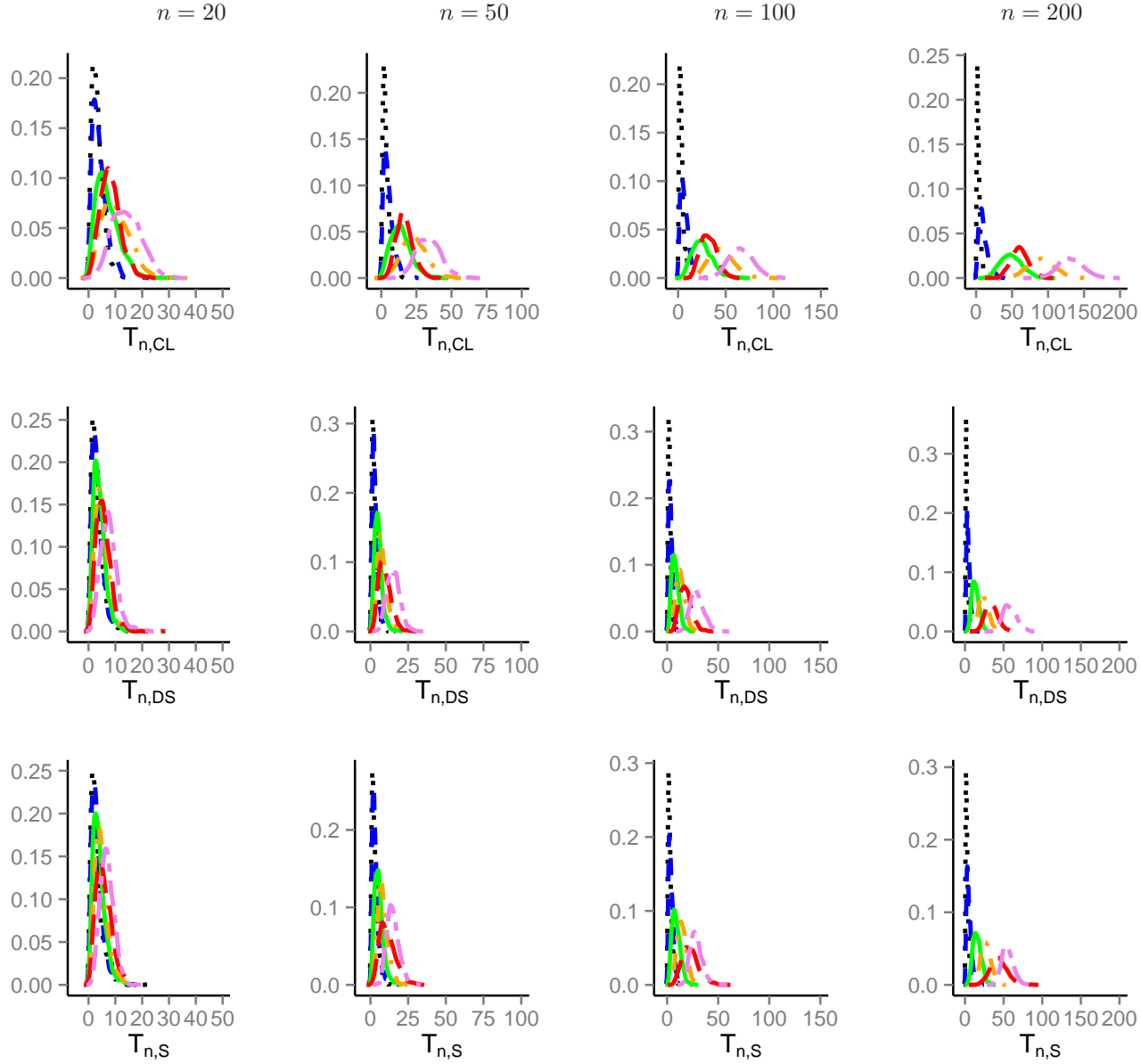


Figure 1: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(1)}$, blue and dash line to $H_{1,0.5}^{(1)}$, green and solid line to $H_{1,1}^{(1)}$, orange and dot-dashed line to $H_{1,1.5}^{(1)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

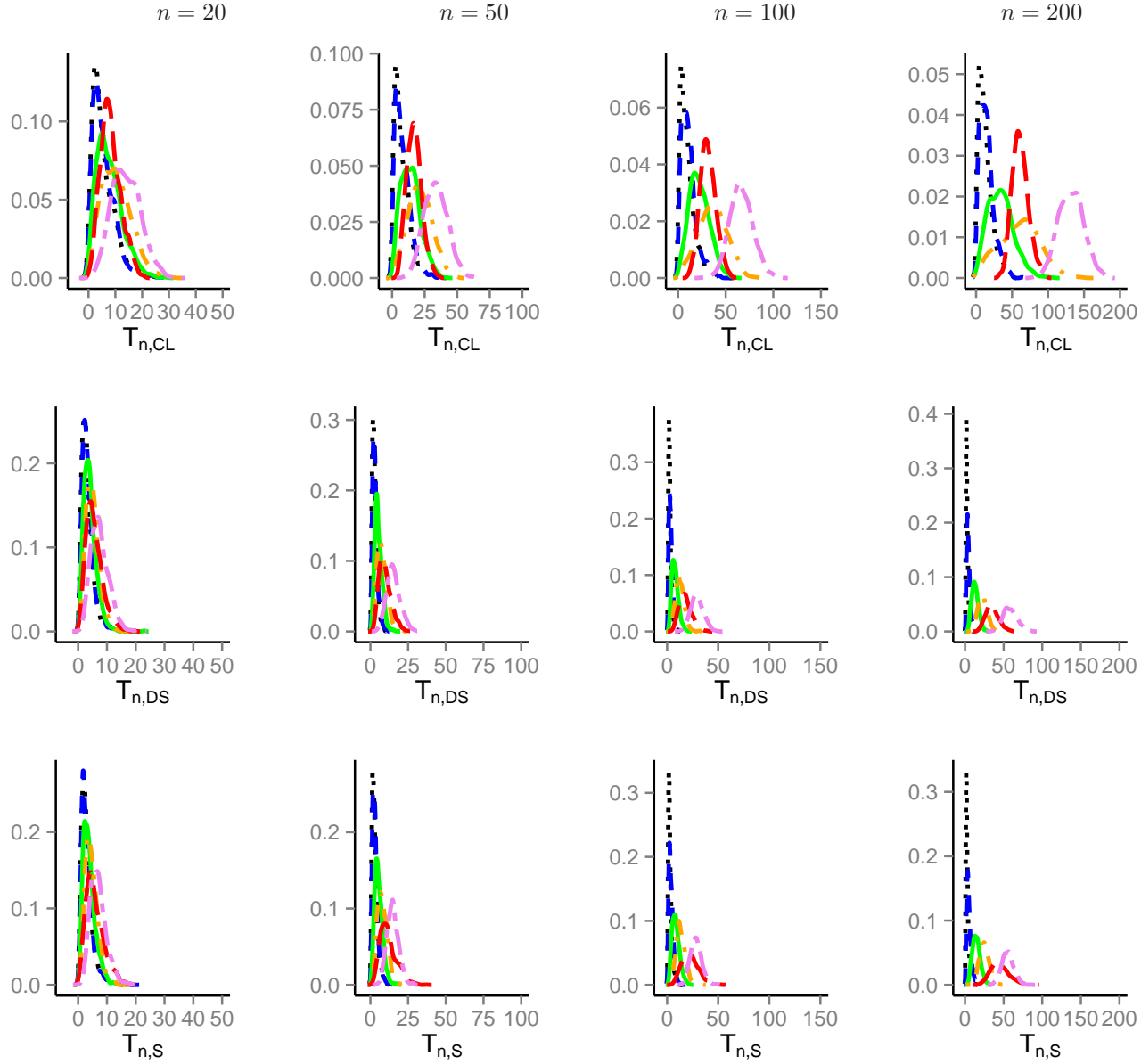


Figure 2: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(2)}$, blue and dash line to $H_{1,0.5}^{(2)}$, green and solid line to $H_{1,1}^{(2)}$, orange and dot-dashed line to $H_{1,1.5}^{(2)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

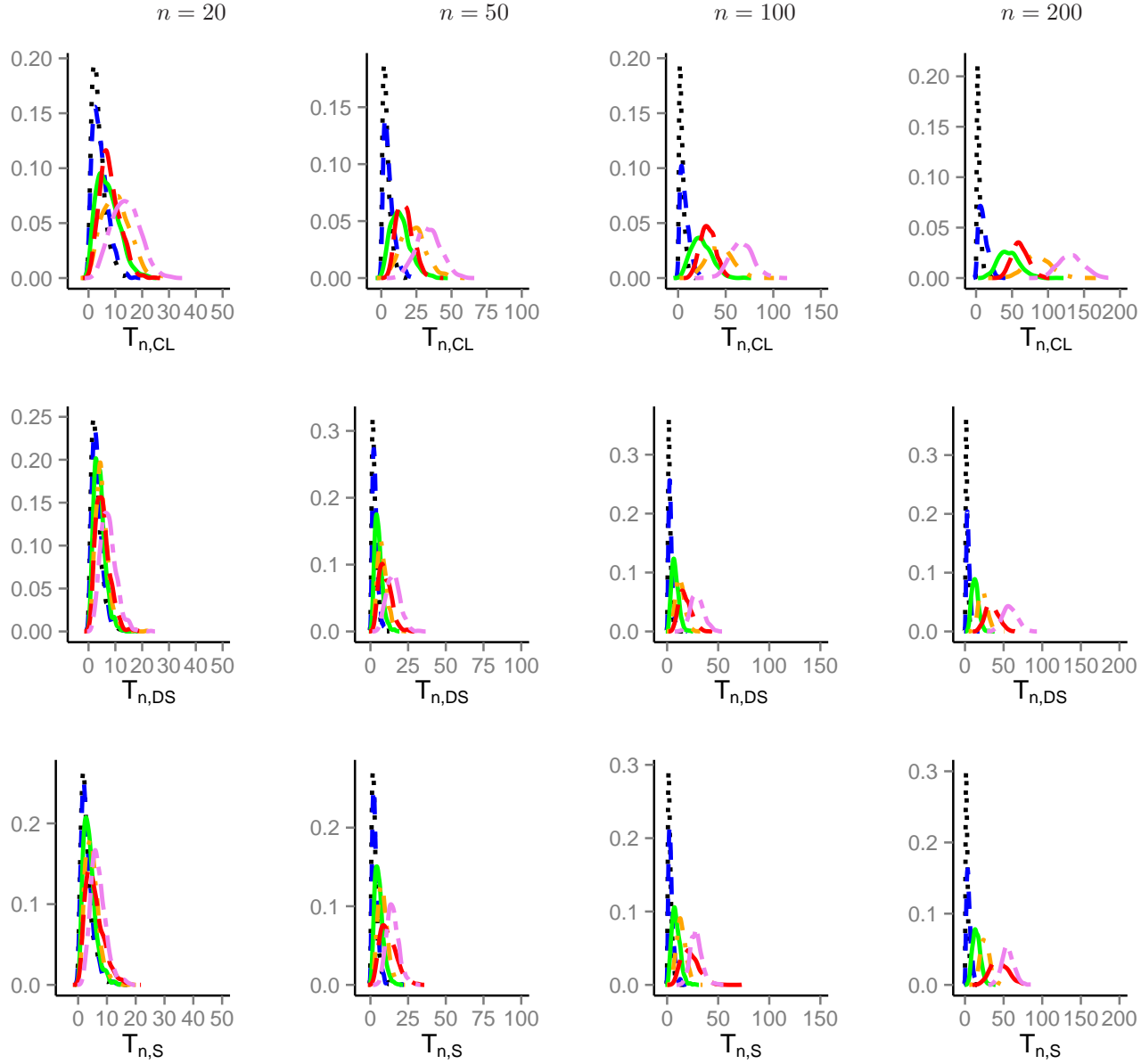


Figure 3: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(3)}$, blue and dash line to $H_{1,0.5}^{(3)}$, green and solid line to $H_{1,1}^{(3)}$, orange and dot-dashed line to $H_{1,1.5}^{(3)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

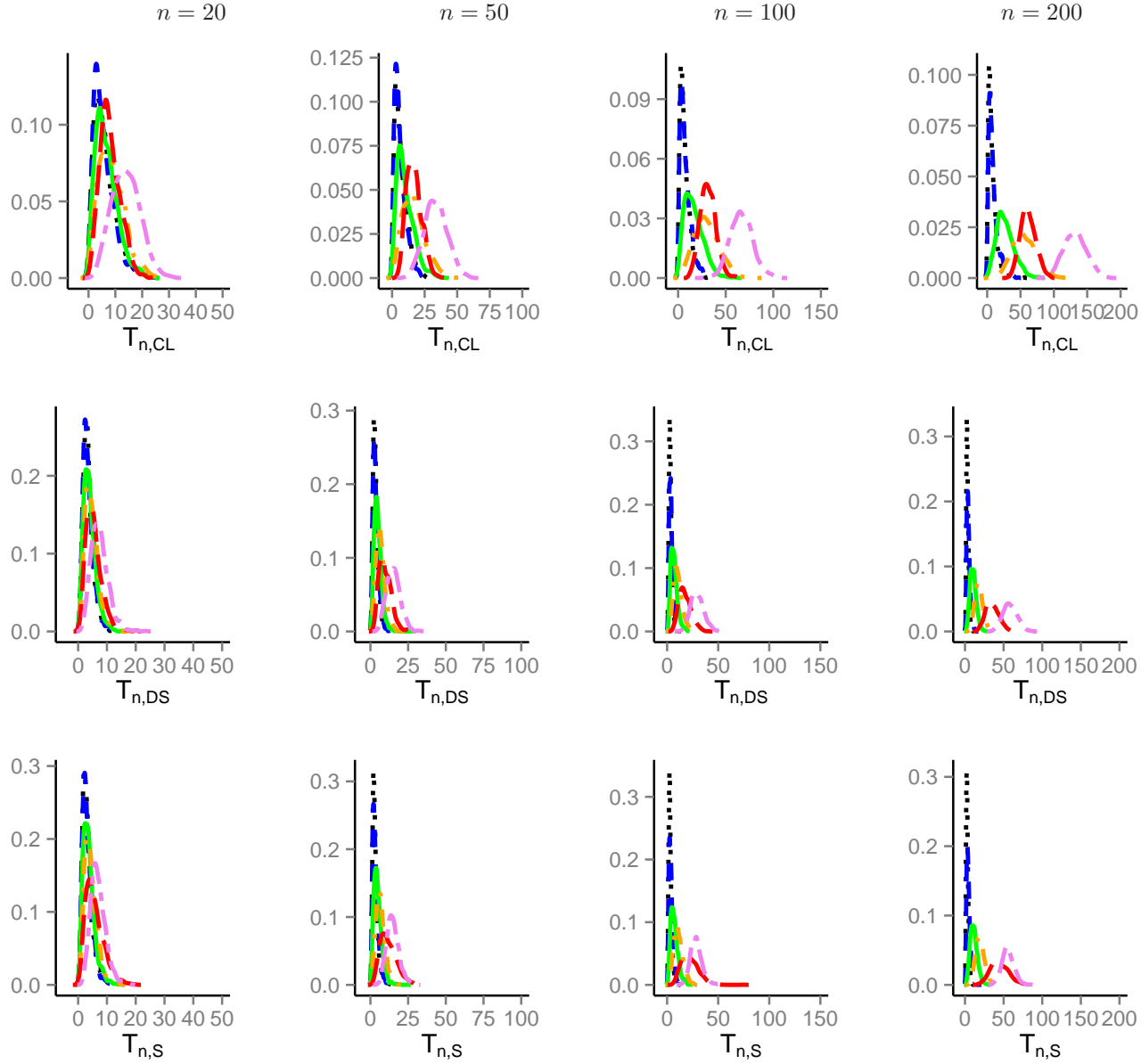


Figure 4: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(4)}$, blue and dash line to $H_{1,0.5}^{(4)}$, green and solid line to $H_{1,1}^{(4)}$, orange and dot-dashed line to $H_{1,1.5}^{(4)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

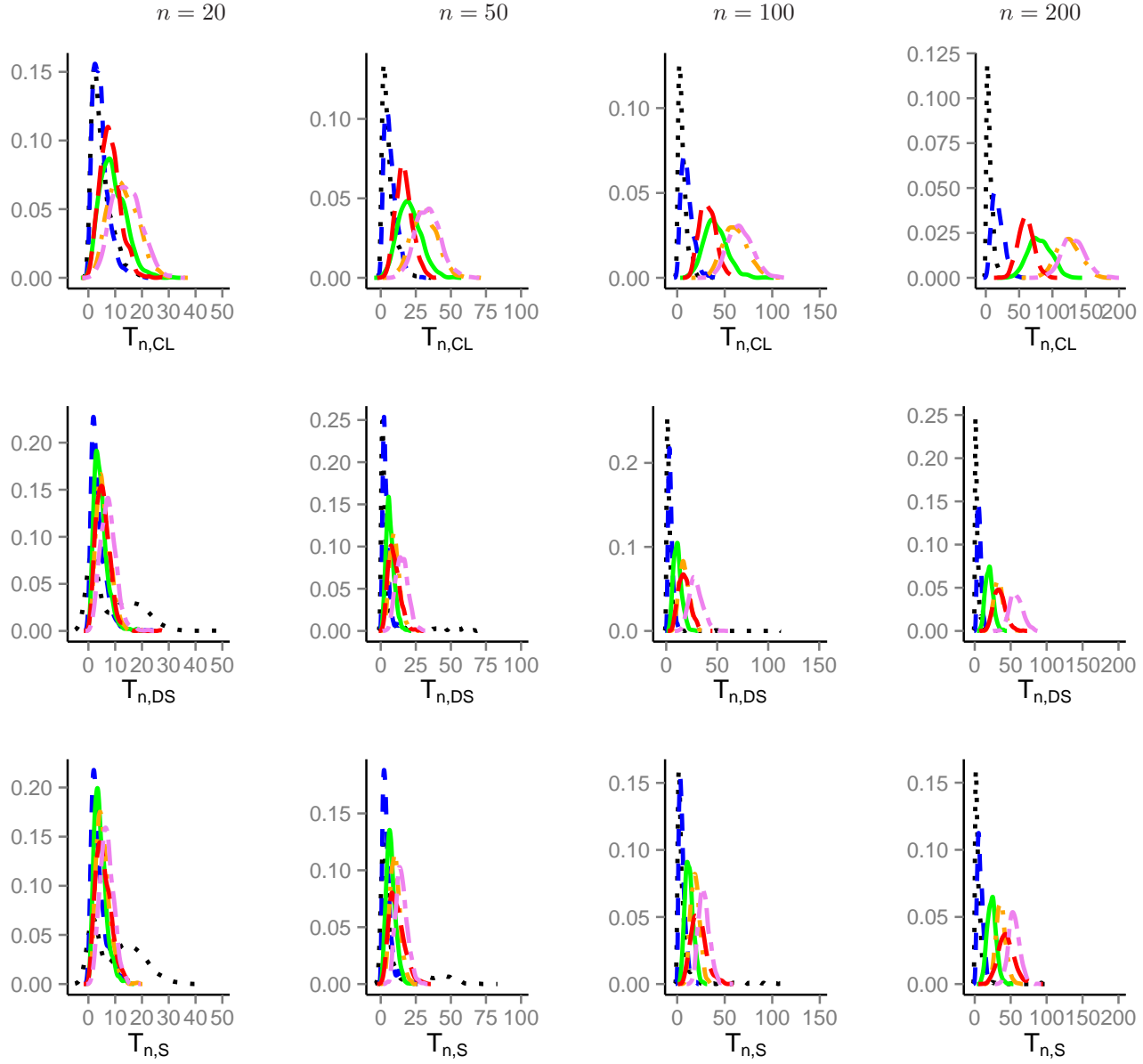


Figure 5: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(5)}$, blue and dash line to $H_{1,0.5}^{(5)}$, green and solid line to $H_{1,1}^{(5)}$, orange and dot-dashed line to $H_{1,1.5}^{(5)}$, red and long-dashed line to $H_{1,\star}^{(1)}$ and violet and two-dashed line to $H_{1,\star}^{(2)}$.

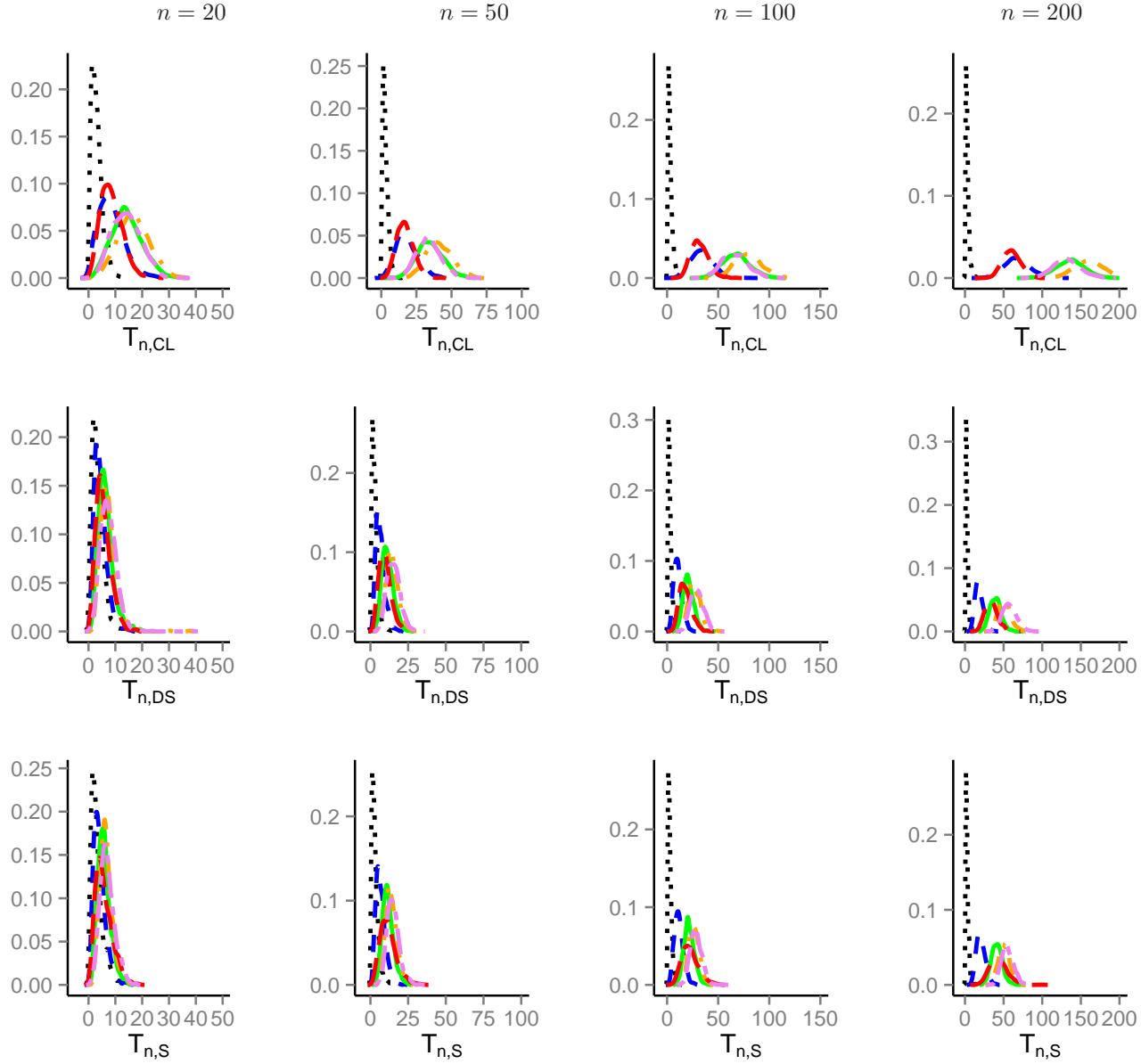


Figure 6: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(6)}$, blue and dash line to $H_{1,0.5}^{(6)}$, green and solid line to $H_{1,1}^{(6)}$, orange and dot-dashed line to $H_{1,1.5}^{(6)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

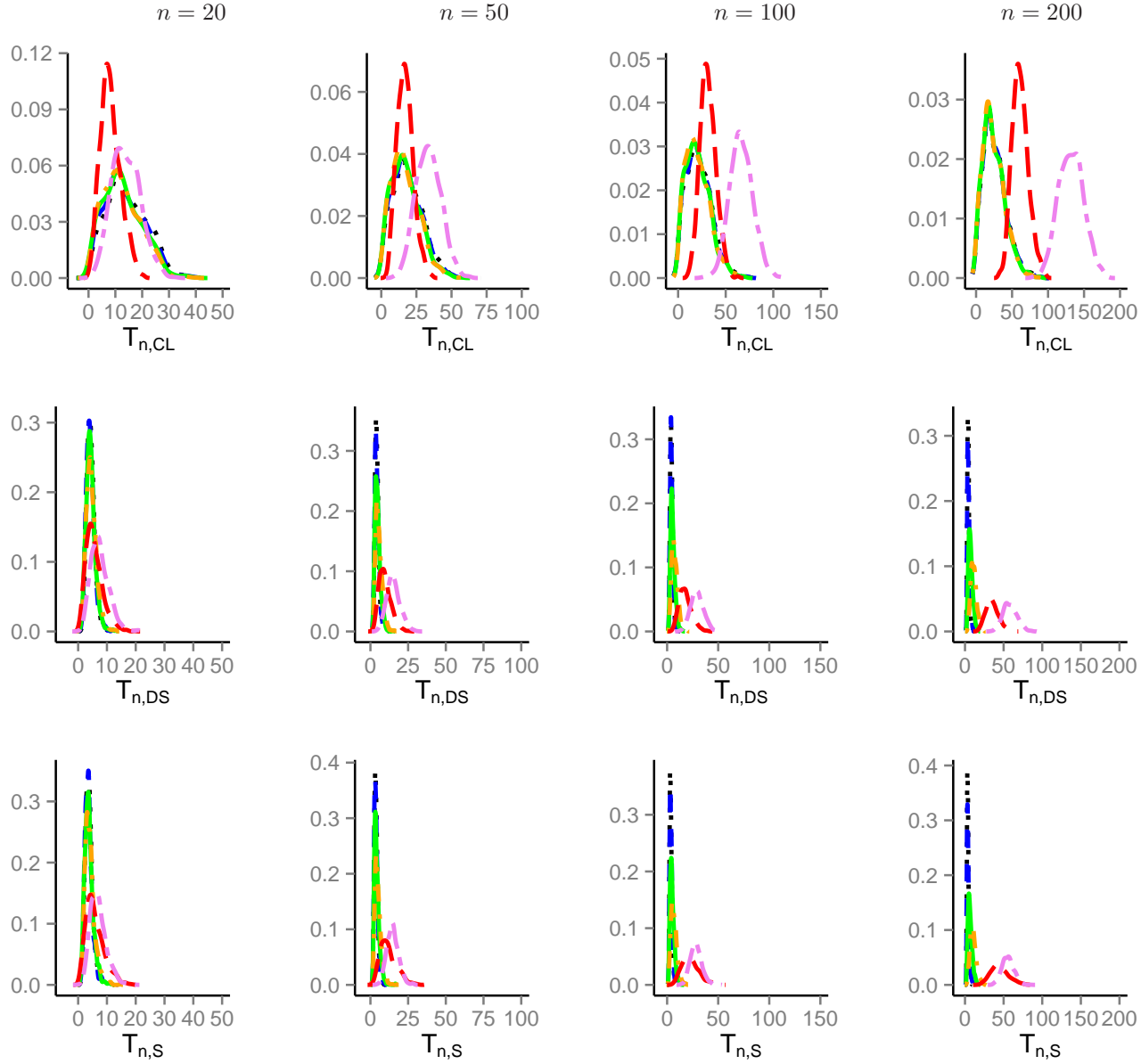


Figure 7: Density estimates of $T_{n,CL}$, $T_{n,DS}$ and $T_{n,S}$. Black and dotted line corresponds to $H_0^{(7)}$, blue and dash line to $H_{1,0.5}^{(7)}$, green and solid line to $H_{1,1}^{(7)}$, orange and dot-dashed line to $H_{1,1.5}^{(7)}$, red and long-dashed line to $H_1^{\star(1)}$ and violet and two-dashed line to $H_1^{\star(2)}$.

References

- Anderson, T. W., Fang, K. T. and Hsu, H., 1986. Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. *Canad. J. Statist.*, **14**, 55–59.
- Baringhaus, L., 1991. Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.*, **19**, 899–917.
- Batsidis, A. and Zografos, K., 2013. A necessary test of fit of specific elliptical distributions based on an estimator of Songs measure. *J. Multivariate Anal.*, **113**, 91–105.
- Batsidis, A., Martin, N., Pardo, L. and Zografos, K., 2014. A necessary power divergence-type family of tests for testing elliptical symmetry. *J. Stat. Comput. Sim.*, **84**, 57–83.
- Beran, R., 1979. Testing for elliptical symmetry of a multivariate density. *Ann. Statist.*, **7**, 150–162.
- Ghosh, S. and Ruymgaart, F.H., 1992. Applications of empirical characteristic functions in some multivariate problem. *Canad. J. Statist.*, **20**, 429–440.
- Fang, K. T. and Anderson, T. W. (eds.), 1990. *Statistical inference in elliptically contoured and related distributions*. Allerton Press, New York.
- Fang, K. T., Kotz, S. and Ng, K. W., 1990. Symmetric multivariate and related distributions. *Monographs on Statistics and Applied Probability*, **36**, Chapman and Hall, London.
- Fang, K.T., Zhu, L.X. and Bentler, P.M., 1993. A necessary test for sphericity of a high-dimensional distribution. *J. Multivariate Anal.*, **44**, 34–55.
- Fernholz, L., 1983. *Von Mises calculus for statistical functionals*. Lecture Notes in Statistics, **19**, Springer Verlag, New York.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.
- Huffer, F. and Park, C., 2007. A test for elliptical symmetry. *J. Multivariate Anal.*, **98**, 256–281.
- Koltchinskii, V. and Li, L., 1998. Testing for spherical symmetry of a multivariate distribution. *J. Multivariate Anal.*, **65**, 228–244.
- Koltchinskii, V. and Sakhanenko, L., 2000. Testing for ellipsoidal symmetry of a multivariate distribution. In: *High Dimensional Probability II*, Eds. Giné, E., Mason, D. and Wellner, J., pp. 493–510.
- Lopuhaä, H., 1989. On the relation between S -estimators and M -estimators of multivariate location and covariance. *Annals of Statistics*, **17**, 1662–1683.
- Morales, D. , Pardo, L. , Pardo, M. C. and Vajda, I., 2004. Rényi statistics for testing composite hypotheses in general exponential models. *Statistics*, **38**, 133–147.
- Muirhead, R. J., 1982. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Canada.
- Schott, J. R., 2002. Testing for elliptical symmetry in covariance-matrix-based analyses. *Statist. Probab. Lett.*, **60**, 395–404.
- Tyler, D., 1982. Radial estimates and the test for sphericity. *Biometrika*, **69**, 429–436.

- Ushakov, Nikolai G., 1999. *Selected Topics in Characteristic Functions*. Series: Modern Probability and Statistics, Walter de Gruyter.
- van der Geer, S., 2000. *Empirical Processes in M -Estimation*. Cambridge University Press.
- van der Vaart, A. and Wellner, J., 1996. *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer.
- Zhu, L.-X. and Neuhaus, G., 2003. Conditional tests for elliptical symmetry. *J. Multivariate Anal.*, **84**, 284–298.